

REALISM, REFERENCE AND THE GROWTH OF SCIENTIFIC KNOWLEDGE

Peter James Smith

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1979

Full metadata for this item is available in
St Andrews Research Repository
at:
<http://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/14770>

This item is protected by original copyright

REALISM, REFERENCE AND THE
GROWTH OF SCIENTIFIC KNOWLEDGE

A thesis submitted to the Faculty of Arts
of the University of St. Andrews by

PETER JAMES SMITH

in fulfilment of the requirements for
the degree of Doctor of Philosophy

September, 1978.



ProQuest Number: 10166519

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10166519

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Th 9253

REALISM, REFERENCE AND THE
GROWTH OF SCIENTIFIC KNOWLEDGE

Abstract

In Chapter 1, I discuss the background to the problems which confront a realist account of the growth of scientific knowledge. At the beginning of Chapter 2, I explain in what sense relativism constitutes a challenge to this account. Four interconnected questions are then posed which are said to underlie the realist position. The chapter finishes with an explanation of how some of them arise in an actual case study. Chapter 3 deals with a general argument of Quine's for the view that reference is inscrutable. In reply I maintain that the argument does not hold good, either with respect to interpreting our own language or when it comes to translating an alien language.

With Chapter 4, I begin to answer the four questions. I explain how Tarski's theory of truth can be seen as a correspondence theory. It is argued, however, that Tarski's theory itself presupposes theories of reference and extension, and certain recent attempts to overcome this lacuna are criticized. In the next chapter I draw an analogy between natural kind predicates, which are of particular importance in science, and proper names, and offer cluster theories of reference for both. This answers the third most basic of the four questions. These theories are defended against criticisms made by Kripke and Putnam. Chapter 6 aims to answer the second most basic question by considering in detail how we can understand what earlier scientific theories were about. I develop some arguments of Davidson's as a counter to Quine's doctrine of the indeterminacy of translation of sentences. I claim that although translation might in fact be indeterminate, Quine,

through concentrating on behavioural evidence to the exclusion of other physical evidence, has failed to show that it is, and that in any case indeterminacy of sentence translation does not imply inscrutability of reference of terms. Finally, in Chapter 7, I sum up my explication of the realist's account of the growth of science with respect to natural kind predicates. I then consider two cases of theory change of different sorts and suggest how my work might be extended.

I, Peter James SMITH, hereby declare that this thesis has been composed by myself, that the work of which it is a record is my own, and that it has not been accepted in any previous application for a higher degree.

In January, 1975, I was admitted under Ordinance General No. 12 as a Research Student of the University of St. Andrews. I subsequently became enrolled for the degree of Doctor of Philosophy. Since January, 1975, I have been a full-time student of the University. My supervisor has been Mr. L.F. Stevenson of the Department of Logic and Metaphysics.

Signed

I hereby certify that the conditions of the Resolution and Regulations for the degree of Doctor of Philosophy in the Faculty of Arts of the University of St. Andrews have been fulfilled in the case of Mr. Peter James SMITH.

Signed

CONTENTS

INTRODUCTION

1

CHAPTER 1 CONFLICTING ACCOUNTS OF THE NATURE AND DEVELOPMENT OF SCIENTIFIC THEORIES

- (i) Logical Positivism: Its Background and
Original Formulation 7
- (ii) Revisions Made to the Original Formulation 17
- (iii) Falsificationism and the Apparent Demise
of Positivism 32
- (iv) A Relativist View of the History of Science 40

CHAPTER 2 A REALISTIC ACCOUNT OF THE GROWTH OF SCIENCE

- (i) Meaning and Reference 47
- (ii) Four Questions for Realism 57
- (iii) Saying What the Extension of a Predicate is 67

CHAPTER 3 THE 'SCRUTABILITY' OF REFERENCE

- (i) Quine on Translation and Meaning 78
- (ii) Evans on Identity and Predication 88
- (iii) Catching gavagai 96

CHAPTER 4 THE FOUNDATIONS OF TRUTH

- (i) Tarski's Theory as a Correspondence Theory 113
- (ii) Truth and Reference 125

CHAPTER 5	CLUSTER THEORIES OF REFERENCE	
(i)	Natural Kind Predicates and Proper Names	136
(ii)	Kripke's Remarks on Naming	147
(iii)	A Theory of Reference for Natural Kind Predicates	162
CHAPTER 6	INTERPRETING PREVIOUS SCIENTIFIC THEORIES	
(i)	Indeterminacy of Translation	179
(ii)	Radical Translation vs. Radical Interpretation	188
(iii)	How Determinate is Translation?	199
CHAPTER 7	A BROADER PERSPECTIVE	
(i)	The Story So Far	218
(ii)	'Phlogiston'	224
(iii)	'Mass'	231
NOTES		240

INTRODUCTION

Philosophers of science have always been exercised by the problems of to what extent scientific theories may be said to describe things that exist in the world and in what sense, if any, scientific theories may be thought of as true. The first problem can be characterized as one of reference. It may be posed as a question — what can we say, on the basis of our scientific theories, about what there is in the world? Or alternatively — what kinds of things is science competent to discover? Some terms which occur in typical scientific theories are 'planet', 'gas', 'bacterium' and 'mammal'. There would be little reluctance to saying that there were things of these kinds in the world. After all, we are mammals that live on a planet, breathe in various gases, and become infected by bacteria. But what about the terms 'gene', 'muon', 'gravitational field' and 'black hole'? Are there also things of these kinds? Perhaps we should be wary here lest they be later regarded as we have now come to regard 'phlogiston', 'caloric', 'luminiferous ether', and the like.

The second problem can be characterized as one of predication. In so distinguishing it from the problem of reference I do not wish to suggest that they are divorced from each other. In fact, the opposite is the case. One way of explaining what it is for any statement to be true is by showing how things in the world can be as it says they are. According to such an account, what it means to say that the statement 'Some bacteria require oxygen to survive' is true, is that there are such things as bacteria and oxygen and that some of the former require the latter if they are to survive. This form of relation between words

and the world is often called a correspondence relation, and a theory of truth based on it a correspondence theory of truth. I shall discuss this theory more fully in Chapter 4.

The position one adopts on the problem of predication, however, need not be so closely related to one's views on reference. One who maintains a coherence theory of truth, for example, according to which a statement is true or false depending on whether it coheres or fails to cohere with a system of other statements, might think that mature scientific theories are true without thereby committing themselves to holding that the world contains the kinds of things the theories postulate.

Philosophy of science in this century has seen a number of distinctive approaches to these problems. Perhaps the most influential have been realism and instrumentalism. A thumb-nail sketch of each might go as follows. The realist maintains that the statements of a theory are either true or false and that most of the things mentioned in a theory of an established science do exist. He therefore adheres to a correspondence theory of truth. The instrumentalist, on the other hand, denies that there are such things, and maintains that while it makes sense to ascribe truth or falsity to reports of observations, it does not make sense to ascribe them to statements of theory. A theory, according to the instrumentalist, figures merely as a rule or principle for the analysis and representation of observed phenomena, and as an instrument in the transition from one set of experimental data to another set.

It is not my intention to discuss in detail in this thesis the relative merits of these conflicting accounts. In Chapter 1, I shall

indicate some shortcomings of both, though it will become clear that I think that those facing instrumentalism are considerably greater and that realism has much more intuitive plausibility. My main interest lies in giving a detailed statement of how a realist explains the growth of scientific knowledge. Here too I think that his account is far more plausible than the instrumentalist's. Both typically talk of a theory's being well-confirmed, of its predictive power, of its simplicity and elegance. The difference is that the realist is in a position to say that our knowledge increases because we learn more and more about the kinds of things there are in the world. Successive theories, says the realist, are in some sense about the same things; they just give a better account of their natures. Since the instrumentalist denies that most of the terms used in scientific theories refer to things in the world, such an account is not open to him. Hilary Putnam puts the point like this,

A natural account of the way in which scientific theories succeed each other — say, the way in which Einstein's Relativity succeeded Newton's Universal Gravitation — is that a partially correct/partially incorrect account of a theoretical object — say, the gravitational field, or the metric structure of space-time, or both — is replaced by a better account of the same object or objects. But if these objects don't really exist at all, then it is a miracle that a theory which speaks of gravitational action at a distance successfully predicts phenomena; it is a miracle that a theory which speaks of curved space-time successfully predicts phenomena; and the fact that the laws of the former theory are derivable 'in the limit' from the laws of the latter theory has no methodological significance.¹

1. H. Putnam, Meaning and the Moral Sciences, p.19.

In recent times, considerations from the historiography of science have led to the realist's views being challenged from another quarter. We have seen the emergence of relativism. According to the relativist there is no one external reality which might be invoked to explain the truth or falsity of statements due to members of linguistic communities more or less distant, in space or time, from our own. He denies the correspondence theory of truth. Nick Jardine describes the conflict as follows,

Each party has sought to convict the other of a distorted interpretation of the succession of scientific theories. Realists charge relativists with commitment to an account which in explaining succession unduly subordinates human rationality to external sociological and ideological factors, and which by denying that it makes sense to talk of cumulative growth of true scientific belief renders the fact of human technological progress inexplicable. Relativists charge realists with 'the chauvinism of time', with commitment to accounts of the contents of past theories and of their succession which are distorted by the imposition of our present conceptual framework and our present criteria for the assessment of theories.¹

Again, it is not my objective to assess all that can be said for and against these contrasting views. At the beginning of Chapter 2, I shall make some criticisms of relativism. Putnam's objections to non-realist views also have to be faced. But I shall not try to somehow rule out relativism on a priori grounds. My concerns are rather to explain realism and defend it against criticism. What I shall attempt to do, then, is to explain how the success of the realist's account of the growth of science is dependent on certain contingent features

1. "'Realistic' Realism and the Progress of Science," Action and Interpretation, ed. C. Hookway and P. Pettit, p.107.

associated with interpreting the theories of previous scientists. This necessitates giving a detailed statement of the realist's view of scientific theories.

At this point I shall take up a number of issues in the philosophy of language. The realist has to be able to explain how we can come to understand that, say, Newton had a theory about gravitational fields, or that Mendel had a theory about genes. He needs to explain how we can interpret previous scientific theories; he needs to give a theory of interpretation. Such a theory, I will argue, presupposes an explanation of what it is for a term like 'planet' or 'gene' to refer. Thus, we need to consider some conceptual problems to do with the notion of reference. Chapters 5 and 6 are devoted to these matters.

The first chapter is mainly historical, looking at the emergence of the three '-isms'. In Chapter 2, I pose four interconnected questions which, I argue, a realist will have to answer if he is to provide a firm foundation for his views. This sets the stage for the rest of the thesis. But before we can start to canvass for prospective answers, it is necessary to be more precise about the correspondence relation. In particular, the realist has to defend his claims about reference against certain pragmatic objections. This is the substance of Chapter 3.

By the time we get to Chapter 7, I hope to have shown that the apparent plausibility of realism is, in fact, well-founded. Through resolving a number of problems in the philosophy of language, a convincing account of the growth of scientific knowledge is secured. The realist, moreover, need not be a chauvinist. He can make sense of the conceptual schemes of earlier scientists, even though he accepts

that by our own scientific lights they were wrong on many points. This is further substantiated by a consideration of two cases of theory change.

For encouraging me in my work for this thesis, and for commenting on drafts of each of the chapters, I should like to thank my supervisor, Mr. Leslie Stevenson. I am also indebted to Dr. Nick Jardine of Darwin College, Cambridge for discussing with me some of the specific problems. Both gave most generously of their time. For providing my main source of financial support during the research I am very grateful to the University of St. Andrews. Finally, I wish to express my deepest gratitude to my wife, Christine, for her constant support and encouragement. In many ways her labours were more than the equal of mine.

CHAPTER 1 · CONFLICTING ACCOUNTS OF THE NATURE AND DEVELOPMENT
OF SCIENTIFIC THEORIES

Section (i): Logical Positivism: Its Background and Original
Formulation

The word 'realism' has been used to connote several different doctrines in the history of philosophy. It is not my intention to survey these various connotations, but it would be useful to look at realism in the broader context of modern philosophy to see how this relates to what I shall call the realist interpretation of scientific theories.

Since the medieval period, realism has generally been understood as the view that material objects exist externally to us and independently of our sense experience. As such it has traditionally been opposed by idealism, the view that no such objects exist apart from our knowledge or awareness of them, and by phenomenism, which denies that material objects exist except as groups or sequences of sense impressions, either actual or possible. Clearly this debate bears some relation to what I characterized in the Introduction as the problem of reference, the problem of what we can say, on the basis of our scientific theories, about what there is in the world. The difference, however, is that whereas the traditional dispute was over the status of things in general, the more recent one in the philosophy of science has tended to focus on theoretical terms as distinct from so-called observation terms. I shall have more to say about this purported distinction shortly. For the present, suffice it to say that theoretical terms are to be thought of as those which typically occur in a theory's

law-like statements, terms like 'electron', 'gravitational field', 'mass' and 'planet', whilst observation terms are to be thought of as those which typically occur in observation reports, terms like 'water', 'red', 'billiard ball', and so on.

As an illustration of the traditional dispute, we might take Berkeley's rejection of Locke's atomistic realism. Locke believed there was a real world of atoms, but that because of their extreme minuteness we are ignorant of their configurations and motions. As an analogous case he cites as contingent our ignorance of very distant bodies.^{1*} Berkeley's challenge to Locke was not so much over whether there are planets or electrons as well as water and billiard balls, but over what the status of any existent thing is. In particular, he was concerned as to the grounds we could have for saying that there was anything more than actual sense impressions. He was arguing against realism tout court rather than questioning whether there were "theoretical entities". More generally we can say that, as regards the period from the time of Descartes until the end of the eighteenth century, it was assumed that the sorts of things that could be discovered by science were of the same kind as those observed in nature. Consequently they were thought to be describable using the same vocabulary and to satisfy the same laws.

The position with respect to what I characterized as the problem of predication, i.e., the problem of whether, and if so in what sense, scientific theories can be correctly described as true, is not quite as clear during this period. Bacon had argued that their truth could not be established a priori but rather had to be established by empirical

* All notes are to be found after the end of the final chapter.

test, by putting questions to nature. Although it was generally realized that, strictly speaking, the verification of predicted phenomena could at best only convey a very high probability on the predicting theory, the question of interest was how, rather than whether, the truth of a theory could be attained. Herein lies one of the main points at issue between the empiricists, following Bacon, and the rationalists, following Descartes. The difference between 'highly confirmed' and 'true' was not debated to nearly the extent that the difference between arriving at scientific theories by inductive generalization and by a priori deductive reasoning was.

Part of the reason for this is undoubtedly the success of both Euclidean geometry and Newtonian mechanics. In the case of the latter, this, together with the lack of any equally good competing theory, suggested that true scientific theories could actually be attained. Naturally the philosophy of the period attempted to accommodate such a realist view of science. This is perhaps clearest when we look at Kant's Critique of Pure Reason, the second edition of which appeared in 1787. Even from the Preface to this edition it is clear that Kant was profoundly impressed by the scope and power of Euclidean geometry and Newtonian mechanics.² This scope and power he attributed to the deductive structure of these disciplines, the deductions being carried out in accordance with the Aristotelian syllogistic. Accordingly Kant sought to establish as synthetic a priori truths not only all arithmetical truths but also the postulates of Euclidean geometry and certain basic principles of physics, such as "in all communication of motion, action and reaction must always be equal"³ — roughly Newton's third law of motion. In characterizing such postulates and principles as a priori, Kant was expressing the view that they were true of

necessity, independently of any empirical knowledge of the world.

Subsequent developments in physical science, however, led to considerable changes in this outlook. Towards the end of the eighteenth century there was a proliferation of physical and chemical theories which made it more and more difficult to relate the postulated structure of matter with the observed phenomena. As a result, unobservable entities like atoms, and unobservable processes like molecular interaction, tended to be regarded rather as heuristic models than as discoveries pertaining to the real world.⁴ Locke's analogy with distant bodies was dissolved. This development was encouraged by situations in which there were too many theories in the field as well as others in which there were too few. As Hesse has commented,

Sometimes there were a number of alternative theories more or less agreeing in their experimental consequences, between which it was very difficult or impracticable to devise crucial experiments. This was the case over long periods with one or two fluid theories of electricity and magnetism, with fluid versus dynamic theories of heat, and with Newtonian force models in chemistry versus Daltonian atomic theories.⁵

By the end of the nineteenth century the assumption that theoretical entities and processes were of the same kind as observational ones was not supported by any chemical or physical theory. The situation became even more acute with the further development of electrodynamics and the emergence of quantum theory early in this century. In particular, the wave/particle duality of light and the discovery by Heisenberg of the uncertainty relations suggested that observable physical objects and non-observable theoretical ones were certainly not describable in the same vocabulary and did not satisfy the same laws. The choice then became one of accepting that theoretical terms refer to real but

non-observable objects which differ in fundamental properties from the observable, or denying that there is anything more than the empirically observable.

This same period saw the doubting of the assumption that science could attain true theories in practice. In place of Newtonian mechanics and Euclidean geometry, whose deductive structures rested on the principles of Aristotelian logic, there arose relativistic and quantum mechanics, and non-Euclidean geometries, whose deductive structures rested on the logic of Frege, Russell and Whitehead. The question became one of whether it even made sense to talk about true scientific theories. Perhaps the only true statements were those involving reference to empirically observable objects, and the most that could be said about a theory would be that it is adequate just in case it contained a certain number of such statements and no false ones.

In the latter part of the last century, Ernst Mach proposed an interpretation of scientific theories which was strongly inclined towards the instrumentalist position on the problems of reference and predication. Initially Mach had held something like a Kantian view according to which every scientific theory is based on certain a priori elements.⁶ Later he rejected this, adopting the phenomenalist view that scientific knowledge consists only of conceptual reflection upon sensory elements. The Laws of Nature, said Mach, are simply "the mnemonic reproduction of facts in thought."⁷ The experiences which they summarize and enable us to anticipate are sensory elements or sensations, like the perception of a colour or a shape, or the feeling of pressure. On this theory we ought not, like the realist, to treat our experiences as experiences of things, nor should we suppose that we have knowledge of things. Harré summarizes it as follows,

Knowledge is only of elements, and considered in relation to ourselves, this is knowledge of the order and sequence of our sensations.

Sensations are the ultimate phenomena and knowledge of sensations the only true scientific knowledge.⁸

Thus, all empirical statements occurring in a scientific theory must be capable of being reduced to statements about sensations. Here we see the emergence of the powerful idea that scientific statements should be empirically verifiable. This later formed the basis for the more general "verifiability theory of meaning", according to which the meaning of a (non-analytic) sentence is in a certain sense identical with its means of verification. For Mach, such means necessarily involved reducing the sentences to ones solely about sense-data.

Early in this century, two groups of philosophers and scientists began to formulate the body of doctrine now generally referred to as logical positivism — the Vienna Circle, under the influence of Moritz Schlick, and the Berlin School, under that of Hans Reichenbach. They readily adopted as a criterion for the meaningfulness of a theoretical statement, i.e., any statement in which a theoretical term is used, that it should be empirically verifiable. With two other aspects of Mach's views, however, they were not so happy. The first was his emphasis on a phenomenalist language which treated sentences about sense-data as basic. Why should this be preferred to a physicalistic language which treated sentences about ordinary physical objects and their properties as basic? The positivists soon came to see these as alternative language forms. The important question was whether the language form was adequate for the needs of empirical description of the world, not whether one language form revealed some ultimate metaphysical truth about the world; the latter they condemned as merely a "pseudo-question".

A further difficulty was that descriptions of sensations could not be made to account for the mathematical relationships underlying scientific laws. With the growth of theoretical physics this became particularly pressing. Moreover, there was the general problem of accounting for the truths of mathematics and logic without introducing unwanted metaphysical notions. As a solution to these difficulties, the positivists turned to the conventionalism of Henri Poincaré. Poincaré came to adopt a conventionalist view through consideration of non-Euclidean geometries.⁹ Gauss, Bolyai and Lobachewski were able to show that, if Euclidean geometry is consistent, there are other consistent geometries containing axioms contrary to Euclid's parallel postulate. Hence, reasoned Poincaré, Kant must be wrong in maintaining that the axioms of a system of pure geometry could be known to be true a priori, since this would mean that their denial would rather be self-contradictory. On the other hand, he argued, geometrical axioms could not be true a posteriori, since they would then be open to continual revision. He therefore concluded that they must be conventions of some kind, "definitions in disguise". Whether Poincaré was right to so conclude is not our concern here. What is of concern is that he also went on to maintain that such theoretical principles as the law of inertia and the law of conservation of energy similarly represented certain agreements or conventions as to how science should talk about phenomena.

Combining the views of Poincaré with those of Mach, the logical positivists proposed the following interpretation of scientific theories. Regularities and irregularities in observed phenomena constitute the subject-matter of science. The former are characterized by the use of theoretical laws which contain theoretical terms. These terms are

explicitly defined via the observation language and are nothing more than conventional abbreviations for descriptions of phenomena. These abbreviations allow for the expression of mathematical relationships, but are in turn merely conventions used in formulating certain relations holding between phenomena.

This interpretation suggests an even-handedness with respect to the problem of reference. Since the meaning of theoretical terms is to be explicated via the observation language, the question of whether there are entities of a different kind corresponding to such terms does not arise. With respect to the problem of predication, though, it was widely held by the positivists early on that empirical statements did admit of conclusive verification. They thought it made sense to talk of the truth or falsity of both observation and theoretical statements, and hence of the truth or falsity of scientific theories.

The final touch which needs to be added to fully characterize their interpretation is that the whole of the theory should be axiomatizable in the first-order predicate calculus with identity. The details of this calculus had been worked out first by Frege. Further work by Russell and Whitehead showed that a substantial part of mathematics could be reduced to it, and so it seemed that the latter-day positivism of Mach and Poincaré could be expressed in the modern idiom, hence deserving the title logical positivism.

It appears that the earliest publication which openly advocated such a view was a 1923 paper by Rudolf Carnap on the task of physics.¹⁰ A more formal version, which draws on later formulations and developments by Carnap, Hempel and others,¹¹ is given below. In particular, the observation language is physicalistic. Throughout the twenties, Carnap not only thought a Machian phenomenalist language

possible, but actually preferred it as a basis for scientific theory. By the mid-thirties, however, its lack of inter-subjectivity, the requirements of mathematics, and various arguments of Popper's, had convinced him otherwise.

According to the early views of the logical positivists, then, scientific theories are to be formulated axiomatically in a first-order calculus with identity, L , in such a way that they satisfy the conditions:

- (i) the non-logical terms of L are divided into two disjoint classes called 'vocabularies':
 - (a) the 'observation vocabulary', V_o , containing observation terms.
 - (b) the 'theoretical vocabulary', V_t , containing theoretical terms.
- (ii) the terms in V_o are interpreted as referring to directly observable physical objects or directly observable properties of physical objects.
- (iii) there is a set of theoretical postulates or basic laws T whose only non-logical terms are from V_t .
- (iv) the terms of V_t are explicitly defined in terms of V_o by 'correspondence rules' C ; i.e., for every term ' F ' in V_t , a definition of the following form must be given: $(x)(Fx \equiv Ox)$, where ' Ox ' is an expression of L whose only non-logical terms are contained in V_o .

Taken together, conditions (i), (ii) and (iv) ensure Mach's criterion for the meaningfulness of theoretical terms.

Obviously several assumptions are made in this formulation: that there is a valid distinction to be drawn between observation and theoretical terms, that we can formalize theories in the way prescribed, that theoretical terms can be given explicit definitions, and that such definitions can take a certain form. Before tackling these assumptions

in section (ii), let us pause to see how the positivists' construal works in practice.

By way of example, suppose that we understand T to be the laws of classical mechanics and C to be correspondence rules which explicitly define the theoretical concepts of T in terms of observable phenomena. Both T and C are formulated in an appropriate logical language.¹² Suppose that we want to apply the theory TC , i.e., the conjunction of T and C , to an experiment involving a solid ball in uniform motion on a flat frictionless plane colliding with a stationary ball. To predict the subsequent behaviour of the two balls, we would first have to determine certain features of the experimental situation, such as the masses of the balls, the velocity of the first, and the angle of impact. These various features can be determined by performing certain simple observational operations, such as noting what numbers the indicator on a balance points to when the balls are weighed. Such observations can be specified in terms of V_o , and then incorporated into correspondence rules which correlate these various observations with terms of V_t in the following manner,

An object x has a mass y iff x is placed on a balance and the

pointer of the balance coincides with the numeral designating number y .

In general, what is happening here is that various observations are being conducted which may be described by certain true sentences of V_o . Using the correspondence rules C these are correlated with various theoretical statements of L such as 'The mass of ball a is b '. These theoretical statements in turn provide a characterization of the experimental situation prior to the collision, and, when taken together with the theoretical laws T , allow us to predict the states of the balls at subsequent times.

It should be clear that the positivists' construal of scientific theories is not intended to be a historical account of how theories actually come to be formed. At the time Carnap and the other positivists were formulating their views, axiomatization was confined to mathematics. A number of theories in the empirical sciences have since been put into axiomatic form; among them, parts of classical and relativistic mechanics, certain segments of biological theory, and the concept of utility in economic theory. But it is still very much the exception rather than the rule. Secondly, correspondence rules are not stated explicitly, even in axiomatized theories; the connection between V_t and V_o is more a matter of implicit understanding. I shall discuss this more at the beginning of the next section. The point I wish to make here is that the positivists' view is intended as an explication of the concept of a scientific theory, not as a description of the practice of scientists when first formulating theories. To the extent to which such an explication is fruitful, it is intended to function as a regulative ideal: the best way to understand a theory is by putting it in this form, for in so doing it is made logically perspicuous and the meaning of any theoretical terms it contains can be understood via the already known observation language.

Section (ii): Revisions Made to the Original Formulation

Over the period of the next forty years, the logical positivists' view underwent considerable modification. Perhaps the most important changes were those relating to condition (iv), which specified the role of the correspondence rules. The rules were said to have the form of explicit definitions which provide necessary and sufficient observational

conditions for the applicability of theoretical terms. In the example, the theoretical term for which I provided a correspondence rule was 'mass'. In his 1936 paper "Testability and Meaning," Carnap drew attention to the fact that dispositional terms such as 'soluble', 'fragile', 'visible' and 'combustible' do not admit of explicit definitions using observational terms, yet are frequently to be found in scientific theories and are clearly meaningful. To see why, let us consider how the term 'combustible' might be defined in terms of observables. An obvious first candidate would be:

A thing x is combustible if and only if it satisfies the following condition: for any time t , if x is placed in contact with a (suitably) hot object at t , then x will burn at t .¹³

An ignited wax taper might, for example, be a suitably hot object. Rendering this definition in first-order predicate calculus, we get:

$$(x)(Cx \equiv (t)(Oxt \supset Bxt))$$

where ' C ' is the theoretical term 'is combustible', ' O ' the observation term 'is placed in contact with a suitably hot object at', and ' B ' the observation term 'burns at'. But this would not give the intended meaning of ' C ', for the universally quantified material conditional on the right-hand side of the biconditional will be true by the ordinary truth-functional definition of ' \supset ' whenever the antecedent, in this case Oxt , is false. That is, ' C ' will be true of anything which is never placed in contact with a (suitably) hot object.

The ordinary material conditional is unsatisfactory. What is required is something which expresses not the idea of what happens if something is placed in contact with a hot object, but what would happen if it were so placed. The form of expression exemplified here is the subjunctive conditional. In order to capture its sense, Carnap

introduced what he called "reduction sentences" and used them to construct more complex "bilateral reduction sentences".¹⁴ The latter were then used to provide partial definitions of theoretical terms. An example of a bilateral reduction sentence which partially defines 'combustible' would be:

$$(x)(t)(Oxt \supset (Cx \equiv Bxt))$$

which may be read as 'if any thing x is placed in contact with a suitably hot object at time t then, if x is combustible, x burns at t, and if x is not combustible it does not'. The one further stipulation made by Carnap is that ' $(x)(t) \sim Oxt$ ' should not be valid, i.e., it should not be the case that everything is such that there is no time at which it has been placed in contact with a suitably hot object. Unlike the attempted explicit definition, if a particular thing a is non-combustible yet has never been placed in contact with a suitably hot object, it does not follow that 'Ca' is true; although ' $(t)(Oat \supset (Ca \equiv Bat))$ ' will be true. The reason why it is avoided is that the bilateral reduction sentence does not completely define what it is for something to be combustible. It rather stipulates a test situation, that the thing be placed in contact with a suitably hot object.¹⁵

Reduction sentences provide only partial definitions of theoretical terms because other test situations could equally well be stipulated. Thus we can say that if a thing is heated, then whether or not it bursts into flame will be an equally good test of its combustibility. Heating in some kind of oven might be imagined here. This suggests the following bilateral reduction sentence:

$$(x)(t)(Hxt \supset (Cxt \equiv Bxt))$$

where 'H' is the observation term 'is heated at'. Here we have an alternative correspondence rule which also may be said to partially

define the theoretical term 'combustible'. Although either correspondence rule constitutes a sufficient condition for the applicability of the term, neither stipulates a necessary condition. All that is required is that the thing would act according to the rule if it were placed in the test situation. Other examples where more than one correspondence rule is appropriate would be: 'explosive' — the thing may be subjected to heat, detonated by another explosion, or have an electric current passed through it; 'fragile' — the thing may be struck, twisted sharply, or subjected to sounds of a high frequency.

A natural reply to this line of argument is to say that, although taken individually the rules do not stipulate necessary conditions, the logical disjunction of them does. But this raises further problems. For one thing, it is doubtful that one could simply state the appropriate set of disjuncts. More importantly though, we need to remember that we are after all talking about a scientific theory which has a definite observation vocabulary used in the specification of its correspondence rules. As the list of disjuncts grows, so does the scope of the theory. It is doubtful that any scientific theory could meet these limitations.

Another reply, this time to the whole debate, would be that it seems like a lot of trouble to go to for the sake of dispositionals — can we not just dispense with them altogether, or at any rate treat them as special cases? Prospects for either alternative are gloomy indeed. Both Popper and Goodman have urged that all properties of things are dispositional.¹⁶ Others have maintained that what dispositionals there are in scientific theories are indispensable.¹⁷ Moreover, we should also note that it is not only dispositionals which imbue scientific theories with a subjunctive character — law-like statements do anyway. To say that lead melts at 327°C is to imply

that, for any ordinary piece of lead, if it were heated to 327°C , then it would melt. Likewise, to say that the children of blue-eyed human parents are blue-eyed is to imply that, if a child were born to such parents, then it would be blue-eyed. The problem of subjunctive conditionals is wider than that of dispositionals. A satisfactory solution has to explain not only how the meaning of dispositional theoretical terms can be specified, but also how scientific laws can, via correspondence rules which relate them to particular instances, support such conditionals.

An interesting special case of the subjunctive is provided by the example at the end of section (i). In the context of the theory of classical mechanics together with the correspondence rule I gave, it is tempting to conclude that what it means to say that the mass of an object x is y is that, if x were to be placed on a balance, then it would have mass y if and only if the pointer on the balance pointed to the numeral y . To do so would be to tread dangerously close to the thin ice of Bridgman's operationalism, according to which concepts like 'mass' and 'length' are synonymous with the statements describing the set of operations used in their determination.¹⁸ Presumably there are other experimental procedures apart from weighing on a balance which may be used to determine an object's mass, although Bridgman thought that here science was confused and that to each procedure there corresponded a different concept.¹⁹ I shall not rehearse the agonies of operationalism here, but rather stand by Hempel's arguments for the partial interpretation of such measurable theoretical terms.²⁰ In consequence, the original correspondence rule needs to be replaced by one such as the following:

If object x is placed on a balance, then the mass of x is y iff the

pointer of the balance coincides with the numeral designating the number y .

In the context of the laws of classical mechanics T and the other correspondence rules, this partially defines the meaning of 'mass' in the sense that it specifies one possible experimental procedure for its determination.

We have seen that the occurrence of dispositional theoretical terms led Carnap to interpret correspondence rules as reduction sentences which define theoretical terms only implicitly. Further reasons for abandoning the goal of explicit definition came to light when we looked at the wider role played by subjunctive conditionals in science. The original condition (iv) cannot be met; therewith the logical positivists were forced to reject the ideal of a synthesis they saw in combining the views of Mach and Poincaré. A theoretical term could no longer be held to be nothing more than a conventional description of phenomena whose meaning consists solely in its means of verification. As Carnap put it, "if by verification is meant a definitive and final establishment of truth, then no (synthetic) sentence is ever verifiable. We can only confirm a sentence more and more."²¹

Let us now look briefly at what later became of reduction sentences and subjunctive conditionals. It might seem that condition (iv) could be replaced by one based on the idea of reduction sentences. It was later argued by Hempel, however, that this is simply not possible for such terms as the $\Psi(x,t)$ function in quantum mechanics, 'electron', 'rigid body', 'point-mass', and measurables like 'force', 'mass' and 'pressure'.²² What was required was a broader view of interpretation, such as that suggested by Norman Campbell's conception of a physical theory as consisting of a "hypothesis", represented by a set of sentences

in theoretical terms, and a "dictionary", which relates the latter to concepts of experimental science.²³ Campbell's dictionary was not of the normal kind, however. It did not so much define theoretical terms as furnish statements to the effect that a theoretical sentence of a certain kind is true if and only if a corresponding observation sentence of a specified kind is true. The dictionary thus provided rules of translation rather than explicit definitions; and partial rules at that, for no claim was made that a translation must be specified for each theoretical sentence.

Suppe summarises the resultant positivist picture of a scientific theory as follows,

The theory ... has various observable consequences which make it testable; but these consequences are not definitional of any particular theoretical terms, being rather the empirical manifestations of theoretical entities interacting in the ways specified by the laws or axioms of the theory. The correspondence rules may be construed as the sum total of admissible experimental procedures for applying the theory to observable phenomena. They do not provide complete definitions of theoretical terms; rather, together with the theoretical postulates T, they provide the theoretical terms with a partial observational interpretation.²⁴

Given this picture, an obvious and important question arises as to what could be meant by saying that a theoretical term has only a "partial interpretation". How does it acquire some of its meaning from the theoretical postulates or basic laws of a theory? This is a difficult question to answer, so difficult in fact that some positivists even chose to forsake the picture which gives rise to it; we shall see why in the next section. Later, in section (iv), we shall also see how this idea came to form, somewhat ironically, an essential part of the more recent criticisms of positivism.

Returning to Carnap's reduction sentences, it can also be shown that they fail to give a complete account of subjunctive conditionals in general. I have already pointed out that a number of reduction sentences could be given for a dispositional property 'P', satisfaction of any of which by an object would be a sufficient condition for 'P' to be truly predicated of that object. It seems inevitable, though, that there will be cases where we recognize an object as possessing 'P', despite its not having been put to any of the tests involved and hence its not satisfying any of the conditions. These will be cases where the disposition is not actually manifested. The reduction sentences will then fail to fully specify what, e.g., 'explosive' means, though we should say we knew perfectly well what it meant — that were the object to satisfy certain conditions, it would explode. These days a more promising account of subjunctives is thought to depend on the use of an intensional logic, much richer than the (extensional) first-order calculus mentioned in my characterization of the positivists' early construal of scientific theories. Clearly, this too calls for some revision of the earlier construal, but to debate it further would take me too far afield.

At the beginning of section (i), I introduced, in an intuitive way, the idea of a dichotomy between theoretical terms and observation terms. Throughout that section I continued to work with the intuitive idea, noted its importance for Mach, and made use of it in formulating the logical positivists' early view. We saw that the positivists came to adopt a physicalistic language rather than a phenomenistic one, but this was shown to have little bearing on the problem of the existence of theoretical entities, for theoretical terms were to be explained, via the verifiability theory of meaning, in terms of the observation

language, whatever it was. In this section we have seen that such a reduction will not generally be possible. Consequently, the meanings of theoretical terms are only partially defined using the correspondence rules to relate them to observable data. They are also supposed to be partially defined in an implicit way by the theoretical postulates or basic laws.

As for the distinction between the observational and the theoretical, the only real attempt at clarification was made by Carnap, again in his paper "Testability and Meaning." Both there and elsewhere he maintained that the distinction was necessarily vague, but chose to make it sharp for the sake of simplicity.²⁵ For the philosopher, said Carnap, the class of observational properties includes ones like 'red', 'hard' and 'hot', while for the physicist it also includes quantitative magnitudes like 'length', 'temperature' and 'time' which can be measured in a straightforward way; for neither does it include 'an electric field of such and such an amount'. Given a predicate 'P' of a language L and an object a, just so long as a person can quickly confirm, on the basis of few observations, using little, but preferably no, apparatus either 'Pa' or '~Pa' to such a high degree that he will either accept or reject 'Pa', 'P' is an observation predicate. Not surprisingly, it was later thought that this explanation, even prefaced by Carnap's qualification, would not do; but it satisfied the positivists until the sixties, and it will do for us until the next section.

We are now in a position to explain more fully the dispute between the realists and the instrumentalists given the positivists' revised construal. Suppose we again take a theory TC and this time consider its set of experimental consequences — the predictions it can make about not just a single experimental set-up like the colliding balls

example in section (i), but about all experiments of the relevant kind. According to the realist, the truth of this set of predictions is a necessary, but not a sufficient, condition for the truth of TC. It is not sufficient because, since the terms of V_t are construed as referring to (possibly non-observable) physical entities, the members of T also have to be empirically true generalizations about those entities. The members of C, in conjunction with T, partially specify the meanings of the terms of V_t and also provide factual statements about the observable manifestations of the entities referred to. The instrumentalist, on the other hand, denies that the terms of V_t refer to anything, and so does not attempt to provide a truth-value for TC. Certainly the experimental consequences will turn out to be true or false, but that is because they contain only terms from V_o . Theories belong to a different linguistic category from that of statements. As Nagel puts it, "theories function as rules or principles in accordance with which empirical materials are analyzed or inferences drawn, rather than as premises from which factual conclusions are deduced."²⁶ The pertinent question for the instrumentalist is whether the theory is adequate in the sense that it enables one to derive just those experimental consequences which are empirically true.²⁷

How adequate are these conflicting views of the cognitive status of scientific theories? The instrumentalist view receives a measure of support from the fact that theories are often formulated in terms of ideal concepts like the physical ones of 'frictionless surface', 'perfect elasticity', 'instantaneous velocity', etc., or the more mathematical ones of 'straight line', 'point', and so on. All that he requires is that the theories in which they occur prove to be effective in representing and implying experimental data. The use of such concepts

might even be cited by way of criticism of the realist, for it would seem difficult to maintain that there are such properties as perfect elasticity and instantaneous velocity, or such things as frictionless surfaces and point masses. In reply it could be said that a property such as perfect elasticity, or objects such as frictionless surfaces, are used as simplifying devices in theory construction and that, when it comes to actually using the theories in practice and collecting data, allowance has to be made for these simplifications. A property like 'instantaneous velocity', on the other hand, functions as a limiting concept. Even though we may not be able to measure it, the fact that a theory in which it occurs provides acceptable data (and why else would we continue to use the theory?) shows the theory to be adequate, and is thus sufficient reason for the realist to ascribe the property to objects. To ~~then~~ reply that measurability is a necessary condition for a property's being "scientifically acceptable" — whatever that may mean — is simply to beg the question against the realist.

A more difficult problem for realism, though again one that instrumentalism can take in its stride, is that apparently incompatible theories are sometimes used in the same experimental domain. Nagel presents it like this,

In inquiries into the thermal properties of a gas we use a theory which analyzes a gas as an aggregation of discrete particles, although when we study acoustic phenomena in connection with gases we employ a theory which represents the gas as a continuous medium. Construed as statements that are either true or false, the two theories are on the face of it mutually incompatible. But construed as techniques or leading principles of inference, the theories are simply different though complementary instruments, each of which is an effective intellectual tool for dealing with a special range of questions.²⁸

He also offers a two-fold reply on behalf of the realist. The first part consists in noting that sometimes one of the theories is more complex than the other but "does not yield conclusions in better agreement with the facts."²⁹ In such a case "the simpler theory can be regarded as in a sense a special case of the more complex one, rather than as a contrary."³⁰ An obvious example here, though not one given by Nagel, is the frequent use of the laws of Newtonian mechanics instead of the more complex, though more accurate, laws of special relativity. Another is that mentioned by Nagel — a particle theory and a liquid or continuous medium theory both used to describe properties of the same substance. The latter is, in a sense, a special case of the former since properties of liquids are those of large numbers of particles acting together; the laws governing aggregates approximate those governing continua "in the large".

The second part of Nagel's reply concerns cases where there appears to be not derivability "in the limit", but genuine inconsistency, like that seen in atomic theory at the turn of this century. Nagel says that here the realist can treat both as temporary for as long as proves necessary; "he can insist on the corrigible character of every theory and refuse to claim final truth for any theory."³¹ For the time being, perhaps, they function as the instrumentalist says — they are instruments used to acquire further experimental data. Meanwhile, as Nagel observes, there is a powerful incentive for the construction of a more inclusive but consistent theoretical structure. He can also point to the reasons for saying that the present theories are inconsistent, i.e., to the reasons for different truth-values being assigned to the same statement. I shall discuss this more at the beginning of Chapter 2.

The realist has his problems, but there does seem to be hope of

solving them. By comparison, the instrumentalist's problems can be shown to be far more serious. An obvious question to put to the instrumentalist is why he should think theoretical terms necessary at all. We have seen that they cannot be explicitly defined using observational terms only, but all that the instrumentalist wants to use them for is deriving the class of experimental consequences from TC which can then be tested. Could this class not be specified straight off, without recourse to TC? The scientist does use theoretical terms, but in a rational reconstruction of his theory is there any reason to? In the past, some instrumentalists thought that an appeal to either Craig's Theorem or the use of so-called Ramsey sentences would justify a negative answer to the latter question. It has since come to be recognized, however, that the correct answer is in the affirmative. Concluding his authoritative study of the debate, Hempel writes, "if it is recognized that a satisfactory theory should provide possibilities also for inductive explanatory and predictive use and that it should achieve systematic economy and heuristic fertility, then it is clear that theoretical formulations cannot be replaced by expressions in terms of observables only."³² As a consequence, instrumentalism seems to be in the rather uncomfortable position of accepting the following theses about theoretical terms:

- (i) they are necessary for the formulation and interpretation of scientific theories,
- (ii) they are neither reducible to nor replaceable by observation terms, yet
- (iii) they do not refer to anything, and
- (iv) they do not mean anything (since it makes no sense to ascribe a truth-value to statements of theory).

Happily it is not the object of this thesis to extricate instrumentalism from this difficulty. In like manner, I shall simply cite some of Popper's criticisms of instrumentalism in the next section without attempting to reply to any of them. What I want to eventually move on to is a problem concerning the growth of scientific knowledge that is posed by a third view of the cognitive status of scientific theories — relativism (or 'super-realism' as it has also been called). In particular, I want to explain how I think a realist should respond to it. But before doing so, I shall end this section by outlining the intuitive ideas about the growth of science which the logical positivists combined with their revised account of the nature of scientific theories. The resultant combination became the primary stalking-horse for the relativist.

Suppose we have some well-confirmed theory TC, i.e., one which explains a large number of observed regularities and has passed a sufficient variety of tests. In the history of science many such theories have come to be replaced by others. How is this to be understood? The positivists' reply was two-fold. In the first place we may extend TC in order to predict new phenomena. This necessitates either the introduction of new correspondence rules or the supplementing of the theoretical postulates, which give the closely related new theories TC' and T'C respectively. We can then test these new theories against their predictions: if they prove incorrect we reject TC' or T'C, although we are free to retain TC; if they prove correct we accept TC' or T'C and we are then free to press on to TC'', etc. Once we have established a theory like TC we can work from it and so increase our stock of scientific knowledge. An example of successful theory extension is the development of the theory of mechanics. Originally it was

formulated to describe the motions of point-masses, and was later extended to encompass the motions of rigid bodies.

Every so often it turns out that, of two theories which were originally formulated in different areas, one may be reduced to, or subsumed under, the other. As examples of this it is common to cite the reduction of thermodynamics to statistical mechanics, the subsumption of the laws of physical optics to quantum mechanics, and the reduction of Kepler's planetary laws and Galileo's terrestrial laws to Newtonian dynamics. This suggests a second way in which well-confirmed theories come to be replaced.

What makes theory reduction much more complex than theory extension is the fact that the theories are from different areas. As Nagel says, "the secondary science [i.e., the theory being reduced or subsumed] employs in its formulations of laws and theories a number of distinctive descriptive predicates that are not included in the basic theoretical terms or in the associated rules of correspondence of the primary science."³³ This means that not all of the V_t terms of the secondary theory are contained in the primary one. 'Heat', 'temperature' and 'entropy', for example, all occur in thermodynamics whereas none of them occur in statistical mechanics. Nagel, in refining the classic treatment of the problem by Kemeny and Oppenheim,³⁴ is thus led to specify the following criteria as necessary for effecting such reductions:

(a) the theoretical terms of both theories must have "meanings unambiguously fixed by codified rules of usage or by established procedures appropriate to each discipline;"³⁵

(b) when a theoretical term 'A' occurs in the secondary theory but not in the primary theory, "(1) Assumptions of some kind must be introduced which postulate suitable relations between whatever is

signified by 'A' and traits represented by theoretical terms already present in the primary science. ... (2) With the help of these additional assumptions, all the laws of the secondary science, including those containing the term 'A', must be logically derivable from the theoretical premises and their associated coordinating definitions in the primary discipline."³⁶

Section (iii): Falsificationism and the Apparent Demise of Positivism

Understanding the background and development of logical positivism has so far been the main concern. Although the positivists' account of the nature of scientific theories dominated discussion in the philosophy of science for much of the early part of this century, it should not be thought that even during this period it was not without its staunch critics who had their own accounts to offer. The staunchest of these, who offered what proved to be a fruitful rival account, was Karl Popper.

From the early days of the Vienna Circle, Popper rejected the verifiability criterion of meaning on the ground that Hume's arguments against the possibility of logically justifying induction showed that scientific theories cannot be verified by any possible accumulation of observational evidence. Indeed, he was convinced that "the problem of meaning" was of no real importance. The positivist attempt to find a criterion of cognitive significance could lead to no positive results, only to the setting-up of arbitrary stipulations, whilst leaving untouched the important problem of distinguishing science from "pseudo-science". "Pseudo-science", according to Popper, includes not only traditional metaphysics but also astrology, which claims to be empirical,

and such ostensibly scientific theories as psycho-analysis.³⁷ As a means of solving this problem of demarcation, he introduced the "criterion of falsifiability": a theory or hypothesis is "scientific" if and only if it can be refuted by observational evidence.

Popper is also strongly critical of those who, like the positivists, seek to analyze theories using artificial logical calculi. The central problem not only in the philosophy of science but also in the whole of epistemology is the growth of scientific knowledge. This cannot be reduced to a study of logical calculi, nor can any such "model language" have a bearing on the problem.³⁸

Not surprisingly, Popper's emphasis on falsification leads him to give a different account of the growth of science. It was noted at the end of the previous section that one way in which the positivists saw science progressing was by the extension of previously well-confirmed theories. But confirmation, according to Popper, is as weak a notion as verification, since it is nearly always possible to find confirmation for a hypothesis or theory. The important question is whether a theory has been thoroughly tested. If so then the extent to which the theory has passed the tests is the extent to which it has been "corroborated". Rightly understood, theories are conjectures, i.e., they are "highly informative guesses about the world which although not verifiable (i.e., capable of being shown true) can be submitted to severe critical tests. They are serious attempts to discover the truth ... even though we do not know, and may perhaps never know, whether [they are] true or not."³⁹ Consequently science ought to encourage the greatest possible proliferation of theories, subjecting a wide variety of theories to possible empirical falsification.

In practice, however, it is never the case that a theory is

rejected simply on the basis of its failing one or more empirical tests. For one theory to be rejected there has to be another which takes its place.⁴⁰ The ideal situation is where we have two theories which both explain a large number of observed regularities but which predict different outcomes given the same experimental situation. In such cases we may appeal to "crucial experiments" to decide between them. Some frequently cited examples of crucial experiments are: the anomalous behaviour of Mercury's perihelion, which was used as a crucial piece of evidence in favour of Einstein's theory and against Newton's; Young's two-slit experiment, which supported Huyghen's wave optics against Newton's semi-corpuscular theory; Michelson and Morley's experiment which was used to discount the theory of the luminiferous ether in favour of the theory of relativity. In each case a previously held, well-corroborated theory was replaced by another.

What Popper is trying to emphasise is that "science is not a system of certain, or well-established, statements; nor is it a system which steadily advances towards a state of finality."⁴¹ The opposite view appears to be that encapsulated in the positivists' account of the growth of science as consisting of the extension of well-confirmed theories and the reduction of one well-confirmed theory to another. Rather than attempting to assess the relative merits of these accounts, though, I shall explain how the relativist's criticisms apply to both.

The emphasis on testability has far-reaching effects; it also leads Popper to reject the instrumentalist view of theories. As we have seen, the instrumentalist views them merely as instruments which enable the scientist to infer further phenomena from phenomena already given. Popper expresses this by saying that the instrumentalist assimilates scientific theories to computation rules or rules of

inference. His criticism of such an assimilation is three-fold: the testing of a scientific theory is carried out in a different way from the trying out of a computation rule; the skill which the application of computation rules requires is quite different from that needed for the theoretical discussion, and the theoretical determination, of the limits of applicability of theories; finally, the theorist's interest is in testable explanatory theories, "applications and predictions interest him only for theoretical reasons — because they may be used as tests of theories."⁴²

According to Popper, the denial of instrumentalism requires acceptance of the view that the scientist aims at finding a true theory or description of the regularities of the world, a theory which will also be an explanation of the observable facts. He further supports the realist by arguing that, although conjectural, theories at least claim to describe something real. Indeed, it is when a theory is falsified that we can say reality has been touched,

Theories are our own inventions, our own ideas; they are not forced upon us, but are our self-made instruments of thought: this has been clearly seen by the idealist. But some of these theories of ours can clash with reality; and when they do, we know that there is a reality; that there is something to remind us of the fact that our ideas may be mistaken. And this is why the realist is right.⁴³

In keeping with this is Popper's "searchlight" theory of knowledge, according to which our discoveries are guided by theory, rather than theories being discoveries due to observation. In the last section I noted the dispositional character of theories and law-like statements. I also noted that Popper held all properties of objects to be dispositional — even 'red' is, he says, since "a thing is red if it is able to reflect a certain kind of light."⁴⁴ He admits that there

are degrees of dispositional character, corresponding fairly closely to the conjectural or hypothetical character of theories, but there is no rigid distinction to be drawn between observation and theoretical terms. All terms, being dispositional, support law-like statements. Hence, "all terms are theoretical to some degree, though some are more theoretical than others."⁴⁵

This naturally leads to an obvious and crucial question: a theory is refuted as a result of observations contrary to those it predicts, but if all terms are theoretical to some degree, what guarantee do we have that an observation report of a crucial experiment will enable us to decide between two different theories? To make this point clearer, suppose we have two competing theories T^1 and T^2 , and a crucial experiment E is proposed to decide between them. The terms used in reporting observations of E will have, according to Popper, some theoretical content; perhaps this content will be derived from T^1 . But in that case why should we even suppose that it could support T^2 rather than T^1 ? We are on the brink of what Feyerabend foresees,

Each theory will possess its own experience, and there will be no overlap between these experiences. Clearly, a crucial experiment is now impossible. It is impossible not because the experimental device would be too complex or expensive, but because there is no universally accepted statement capable of expressing whatever emerges from observation.⁴⁶

Popper's argument against the distinction between observation and theoretical terms can be avoided if, as Carnap held, theoretical terms can be distinguished without recourse to dispositionals. But there is a problem with Carnap's account even at the outset. It will be remembered that he distinguishes between those properties considered observational by the philosopher and those considered observational

by the physicist. This already suggests a four-fold distinction. Generalizing this point, Achinstein writes, "how many aspects of an item, and which ones, I must attend to before I can be said to observe it will depend upon my concerns and knowledge."⁴⁷ It begins to sound as though what counts as an observation varies from observer to observer with the amount of knowledge each brings with them.

N.R. Hanson coined a very useful word here: he urges that what we ordinarily consider to be descriptive, observation terms are "theory-laden".⁴⁸ That is, we apply them specifically in the light of the theories and laws that we happen to subscribe to. As an example he considers a physicist looking at a piece of apparatus and seeing an X-ray tube. The neophyte, looking at the same object, observes rather "a glass and metal instrument replete with wires, reflectors, screws, lamps, and pushbuttons."⁴⁹ Some have interpreted Hanson as saying that the theories we use in describing the world themselves determine what objects there are and what properties they have. But one set of objects does not suddenly disappear and another set materialize simply because we change our theory! Properly construed, what is being questioned is the extent to which both the kinds of objects there are observed to be and the properties they are observed to have are dependent on the theories presumed by the observer.

A further weakness with Carnap's attempt to draw the distinction is that there are frequently cases where we can decide, on the basis of few observations, unaided by apparatus, that a sentence 'Pa' is true, even though 'P' would ordinarily be regarded as a theoretical term. We often decide changes in kinetic energy, mass and electric charge in this way. (Imagine your hair standing on end as you approach an operating Van der Graff accelerator — have you observed an electric

field?) More generally, how many observations are to be required, and what are they observations of? A physicist quickly identifies a track in a cloud chamber as that of a decaying β -particle; an artist examines a blob of paint for some time and still cannot decide whether it is aqua or ultramarine.

Strong criticism of the distinction also comes from Putnam.⁵⁰ To begin with, he claims that "writers like Carnap must be neglecting the fact that all terms — including the 'observation terms' — have at least the possibility of applying to unobservables."⁵¹ In support of this he points out that 'red' was so used by Newton when he postulated that red light consists of red corpuscles. Putnam also notes that observation statements may well contain theoretical terms and gives as an example the statement 'We also observed the creation of two electron-positron pairs'. This latter claim receives support from Hanson's argument noted above.

Such arguments proved conclusive against the observation/theoretical distinction as traditionally conceived by Carnap and the other positivists. They do not show that no distinction of the kind can be drawn, but one has to question the point of having a distinction anyway; what do we want it for? We have uncovered one possible reason — to provide a neutral language in which to compare rival theories using crucial experiments. This need, however, arose through the consideration of Popper's methodology; what did the positivists want it for? The answer takes us back to a doctrine of Mach's. The primary rationale for the distinction was to provide an empiricist methodology, and it aimed to do so by showing how the sentences of a language suitable for the expression of a formalized scientific theory were cognitively significant. The cognitive significance of those non-analytic sentences

containing terms from the theoretical vocabulary V_t of a theory was supposed to be ensured by their partial interpretation using the correspondence rules of a Campbellian-style dictionary. As there is no clear way of distinguishing V_t from V_o for a theory TC, a fortiori no clear sense can be given to the task of specifying the meanings of the members of the former class using members drawn from the latter.

Two further points need to be added. As a consequence of its reliance on a clear observation/theoretical distinction, Carnap's idea of partial interpretation, according to which part of the meanings of theoretical terms is given through correspondence rules and part through the theoretical postulates or basic laws, came to be abandoned. As with the distinction itself, the arguments against partial interpretation do not show that no content can be given to the notion that the meaning of a term like 'electron' is specified, in part, by the theory of electrons.⁵² Indeed, the notion does seem to contain an element of truth, and this lends support to the criticisms of Hanson and others. I shall discuss this further in the next chapter. As a point of historical fact, though, the common reaction amongst the positivists was that of Hempel who concluded that, whichever way one looks at it, the notion of partial interpretation is incorrect as a picture of the way in which a scientific theory works; consequently, "the doctrine that the meanings of theoretical terms are implicitly specified, at least in part, by the theoretical calculus must therefore be rejected."⁵³

The second point is that the very existence of a distinction between what is and what is not cognitively significant was shown to be what Quine termed a "dogma of empiricism".⁵⁴ We saw in the first section that the verification criterion of meaning was one attempt at specifying the cognitively significant sentences which were not analytic,

i.e., those which were synthetic. The change from verification to confirmation and the attendant loosening of the requirements imposed on the correspondence rules parallels a loosening in the positivists' criterion of cognitive significance and hence of which sentences are to be regarded as synthetic. But either way, a specification of cognitive significance amounts to nothing more than the requirement that every sentence of a theory TC should be either analytic or synthetic and not both. Thus cognitive significance and the analytic/synthetic distinction both serve the same purpose. As Quine puts it, "the two dogmas are, indeed, at root identical."⁵⁵ To the extent to which the dogma is discredited, so the observation/theoretical distinction underlying it is ill-motivated.

Section (iv): A Relativist View of the History of Science

A further radical challenge to logical positivism was over the account it adopted of the growth of scientific knowledge. Towards the end of section (ii), it was pointed out that Hempel, Oppenheim and Nagel embraced the ideas of theory extension and theory reduction. We have been given good reason for believing that this is not the whole story; Popper's falsificationist account needs to be recognised too. But surely, Popper's criticisms notwithstanding, it is at least part of the story. A scientific theory, be it well-confirmed or well-corroborated, can sometimes be extended to other domains in which it has so far not been applied or tested. Likewise it is undoubtedly the case that we do have apparent instances of one well-confirmed or well-corroborated theory which is reducible to another. The realist can give a plausible explanation both of what makes such progress possible

and of how it comes about. His basic claim is that theories of a mature science describe the behaviour of things in the world. Scientists act in accordance with this view in attempting to provide theories which better explain the behaviour of the same kinds of things dealt with by earlier theories. The progress of science is evidence that such a methodology works. Thus, the realist wants to say such things as that what Bohr identified as electrons earlier in this century, modern physicists do too, that most of the things Dalton classified as acids are similarly classified by our chemists, that temperature is nothing more than the measure of the mean kinetic energy of molecules, and that the planets Kepler recognised through Tycho Brahe's observations and to which he applied his laws of motion are just those to which Newton applied his laws of dynamics. Popper, as we have seen, is sympathetic to the realist claim. Yet both positivist and falsificationist accounts, and in consequence the realist claim, have been placed in doubt. We saw this in part with Feyerabend's challenge to falsificationism. I now want to move onto his criticisms of positivism.

At the end of section (ii), I noted the conditions Nagel specifies for the reduction of one theory to another. The second part of condition (b) was that all the theoretical and experimental laws of the reduced science should be logically deducible from the theoretical laws and correspondence rules of the reducing science. Feyerabend has questioned whether this condition, which he usually calls the condition of "consistency", is ever fulfilled.⁵⁶ He has also challenged a second condition, which he says is "an immediate consequence" of the first. It is that "the meanings of the primitive descriptive terms of the secondary science ... will not be affected by the process of reduction."⁵⁷ This he calls the condition of "meaning invariance", and he interprets

Nagel's condition (a) as a statement of it. Taken together, what Feyerabend is calling into question is the whole positivist account of the growth of science.

Feyerabend does not direct his criticisms at the positivist conditions as they relate to empirical generalizations of the 'All-ravens-are-black' type, "which abound in the more pedestrian parts of the scientific enterprise," but to "universal theories" like "the Aristotelian theory of motion, the impetus theory, Newton's celestial mechanics, Maxwell's electrodynamics, the theory of relativity, and the quantum theory."⁵⁸ The conditions, he claims, are inapplicable to advancements involving such genuine scientific theories. In an attempt to demonstrate this he gives a series of case studies of purported reductions which, he argues, show the conditions to be violated.

Let us briefly review some of his examples concerning the consistency or deducibility condition. Consider the statement that Galileo's physics is reducible to the physics of Newton, an example commonly cited by Nagel and others. Part of what is meant by this statement is that the laws of the former may be logically deduced from the latter. Consequently Galileo's theory of the motion of objects near the surface of the earth should be deducible from Newton's laws of dynamics. A basic assumption of Galileo's theory is that vertical accelerations in free fall near the earth's surface are constant over any finite (vertical) interval. Given Newton's theory, however, vertical accelerations in free fall are inversely proportional to distance from the earth. Admittedly, the difference may be experimentally indistinguishable, but that has no bearing on the fact that, strictly speaking, the two theories are logically inconsistent.⁵⁹ As regards further cases,

Feyerabend remarks, "that statistical thermodynamics is inconsistent with the second law of the phenomenological theory; that wave optics is inconsistent with geometrical optics; and so on."⁶⁰

Turning to the condition of meaning invariance, Feyerabend takes as a paradigm example the reduction of classical mechanics to relativity theory.⁶¹ The classical theory assumes that the mass of a particle is constant and is conserved in all reactions in a closed system.

According to relativity theory, however, the mass of a particle is proportional to its velocity relative to a co-ordinate system in which the observations are carried out. To appreciate that what is at issue here is a change in the meaning of the term 'mass', one has to examine the structures of the theories and the roles played by the term in both. In the first place, different and (apparently) incompatible equations about mass hold in the two theories. Secondly, relativistic mass is a relation, involving relative velocities, between an object and a co-ordinate system, whereas classical mass is a property of the object itself and independent of its behaviour in co-ordinate systems. Nor will it do, he adds, "to identify the classical mass with the relativistic rest mass ... for although both may have the same numerical value, they cannot be represented by the same concept."⁶²

In the same context, Feyerabend considers Nagel's condition (b)(1), in which assumptions of some kind are postulated in order to establish relations between concepts of the two theories. Presumably this would amount to recognising that "under certain conditions the occurrence of relativistic mass of a given magnitude is accompanied by the occurrence of classical mass of a corresponding magnitude."⁶³ Yet this too is inconsistent with relativity theory which asserts that there are no absolute (classical) masses and hence that mass, as understood in

classical physics, does not express an actual property of a physical object.

Other examples treated at length by Feyerabend concern the concept of impetus as it occurs in Aristotle's theory of motion and in Newton's mechanics, and the concept of temperature as it occurs in phenomenological thermodynamics and as it might be represented in statistical mechanics. In all these cases, the conclusions he arrives at result from his taking seriously the view of Nagel and the later positivists that the meaning of a term depends upon the theoretical context in which it occurs. That is, his arguments rely on the point stressed by Nagel in condition (a) and elsewhere that, "it is ... of utmost importance to note that expressions belonging to a science possess meanings that are fixed by its own procedures of explication."⁶⁴ We should also recall here that his conclusion of the impossibility of conducting crucial experiments resulted from following through with Popper's adherence to a similar thesis about the theoretical content of terms used in observation reports. The radical conclusion which appears to follow is that rival or successive theories may use the same terms but in principle there is no way of deciding whether or not they mean the same things by them. Feyerabend and Kuhn describe such a situation as being one in which the theories concerned are incommensurable.⁶⁵

What now becomes of the realist claim that rival or successive theories can be about the same things? Feyerabend defends the view that the ontology or domain of an earlier theory is completely replaced by that of a subsequent one.⁶⁶ Strictly speaking, the only entities there are, are the ones contained in the ontologies of the theories we currently accept. To quote Feyerabend, "introducing a new theory involves changes of outlook both with respect to the observable and

with respect to the unobservable features of the world, and corresponding changes in the meanings of even the most 'fundamental' terms of the language employed."⁶⁷

On the basis of a rather different analysis of the history of science, Kuhn reaches a similar view with respect to theories separated by a 'scientific revolution'.⁶⁸ An old scientific "paradigm" is occasionally displaced by a new one, with the result that, in some senses at least, the post-revolutionary scientist finds himself working in a "different world".⁶⁹ Other philosophers to have expressed general agreement with this line of argument include Sellars⁷⁰ and Maxwell.⁷¹ Unlike the instrumentalists, they do not dispute the realists' view that scientific theories are true descriptions of what there is. By conjoining this view with the positivist thesis that the meaning of a term is to be explicated via the theory (or theories) in which it occurs, they rather take realism to an extreme whereby 'what there is' depends on what theory or paradigm is in use. Thus Mellor has aptly termed their position "super-realism".⁷² Others see it as expressing a form of relativism.

This is taking things rather too quickly though. Feyerabend's suggested counter-examples raise serious doubts about Nagel's conditions. The incomparability of meanings for terms in different theories does seem to follow from the view that the meaning of a scientific term depends on the theory in which it occurs. But history should teach us something here. The positivists never did give a full account of what the meaning of a term is, and not surprisingly they found it difficult to articulate the idea that the meaning of a theoretical term is partially specified by theoretical postulates. This should prompt us to look more closely at the rather glib pronouncement 'the meaning

of a scientific term depends on the theory in which it occurs'.

A problem in the philosophy of science leads to an inquiry into the nature of meaning — a problem in the philosophy of language. I shall take this up at the beginning of the next chapter. To end this one, we should note one further important point.

Suppose that Feyerabend is right about the incomparability of meanings between terms of different theories. This fails to show that we cannot decide the weaker thesis of whether or not rival or successive theories are about the same things. It might not be determinate exactly what, say, earlier theorists believed about those things, but this is a different matter. The realist account of the growth of scientific knowledge relies, in the first instance at least, only on the weaker thesis.

CHAPTER 2 A REALISTIC ACCOUNT OF THE GROWTH OF SCIENCE

Section (i): Meaning and Reference

Let us begin by taking a closer look at the statement 'the meaning of a scientific term depends on the theory in which it occurs'. In talking thus of 'a scientific term' I presuppose no rigid distinction between terms that are and terms that are not in some sense observational. As we have seen, any distinction of this kind is at best ill-motivated and at worst untenable. A loose distinction between different kinds of statement will, however, sometimes be assumed. This will enable us to distinguish obvious statements of theory, like 'the gravitational force between two bodies is directly proportional to the product of their masses and inversely proportional to the square of the distance between them', from what are clearly observation reports, like 'object x rotated in a clockwise direction around object y'.

An obvious first question to raise about the statement under consideration is whether all of the principles used in stating the theory serve to specify the meanings of all of the terms contained in it, or whether the relation is a somewhat weaker one. To start with, let us concentrate on the former, stronger alternative. Some support for such a view might come from a "network" picture of the meanings of scientific terms: 'P¹' and 'P²' occur in theoretical principle 'Th¹', 'P²' and 'P³' in 'Th²', and so on; 'Th¹' does not contribute directly to the meaning of 'P³', but only indirectly in virtue of its contributing to the meaning of 'P²' and the latter's occurring in 'Th²'.

One consequence of this alternative is that a change in any of a theory's principles leads to a change in the meanings of all of the

terms contained in the theory. Feyerabend, who at one time seems to have held the strong position, finds this consequence unpalatable.¹ As an extreme example, we might consider his case involving two theories of classical celestial mechanics, differing only in that one gives a slightly different value for the strength of the gravitation potential. Feyerabend rightly concludes that it would be rash to maintain that a transition from one theory to the other would somehow involve a change of meaning in the terms.²

There is a lot of play here between change of theory on the one hand and change of meaning on the other. Whilst it does seem clear from the above example that minor changes in numerical values for constants does not lead to changes of meaning, it might be replied that in so far as this is the only difference we have not changed from one theory to another but merely slightly altered the first. Apparently what is required is a better grip on the notions of change of theory and change of meaning. Before discussing them further though, we should note some other problems with the strong alternative, ones that also cast doubt on the tenability of Feyerabend's position as it was outlined at the end of the previous chapter.

In discussing the thesis of meaning variance, Feyerabend holds that apparently successive or rival theories may contain incompatible statements. But by maintaining that all the terms of such theories are incommensurable, i.e., that there is in principle no way of showing that they mean the same, it becomes obscure how certain of the statements contained in the theories could be shown to be incompatible.³ If there is no way of telling whether by 'mass' a Newtonian means the same as an Einsteinian, then when the former assents to the sentence 'Mass is a

constant property of an object' and the latter dissents from it there is no way of telling whether they are even assenting to and dissenting from the same statement; they are talking past each other. Showing that statements from different theories are incompatible appears to involve at least the possibility of translating from one to the other. The possibility of translation, however, is precisely what incommensurability denies.

A further problem is revealed when we press the point about there being no theory-neutral observation language. In the last quotation of the previous chapter, Feyerabend talks about "changes of outlook both with respect to the observable and with respect to the unobservable features of the world." But what content can be given to the idea of such changes where all of the terms of the theories are incommensurable? A similar difficulty arises with his comments that Galileo's laws and Newton's theory conflict in a common domain of validity,⁴ and that a wider, more embracing theory is capable of covering all the phenomena covered by a narrower theory which it succeeds.⁵ If Feyerabend's radical conclusions stand and all of the principles of a theory serve to specify the meanings of all the scientific terms, it makes no sense to talk of common domains of validity or of the same sets of phenomena forming the subject-matter of incommensurable theories.

As a final point, let us consider what account of the testing of theories is open to the holder of the strong alternative of the thesis about meaning. Suppose we have some theory T which predicts an observation O. Could it be the case that when the observation is carried out the resultant report entails not-O? Ordinarily we should certainly think so — not even the most successful theories are necessarily true,

i.e., are such that there could be no possibility of an observation contrary to that predicted. Since there is no language independent of theory for the reporting of the observation, however, terms employed in describing predictions of T will be theory-laden and statements of predictions will themselves contribute to specifying the meanings of other terms in T. It follows that if not-O were to be accepted, the meanings of the terms used in its reporting could not be assumed to have remained the same as those used in the original prediction O. By way of example, consider one theory of celestial mechanics — Galileo's for instance — which predicts that a planet's orbit is circular. Observation of the solar system suggests rather that the orbits are not circular but more nearly elliptical. Yet if observation reports to this effect were accepted, the meaning of 'orbit of a planet', and several other terms no doubt, could not be said to be the same for both theory and observation. Hence no observation report is possible which could disconfirm or falsify the theory. That is, the only observation reports which are relevant to testing the theory will be those which are consistent with the theory. All testing of theories is thus circular.

Evidently there are serious problems with the view that all of the principles used in stating a theory serve to specify the meanings of all of the terms contained in it. Paying strict attention to this view robs it of those considerations which originally gave rise to it. Let us move on, then, to the question of how we might weaken it in order to avoid these paradoxical consequences.

One alternative would be to hold that only some of the principles of a theory serve to specify meanings. Other alternatives are suggested if it is held that theoretical principles only contribute to meaning

specification, i.e., they only partially specify meanings. In the previous chapter, though, we saw that each of these alternatives engenders difficulties. The first rests on the idea that we can distinguish between "truths grounded in meanings" and "truths grounded in facts". But this is precisely one of the dogmas attacked by Quine in his famous paper.⁶ As for the notion of partial specification of meaning, the difficulty remains of showing how this can be freed from the assumption of a firm distinction between observation and theoretical terms, thereby divorcing it from Carnap's ill-fated notion of "partial interpretation".⁷ So while there does seem to be something intuitively correct about the idea that the meaning of a scientific term depends on the theory in which it occurs, it has so far resisted all attempts at clarification.

What we have here is a symptom of a much more widespread complaint. For there to be a theory of meaning for any given class of terms, such as those used in stating scientific theories, it would seem that there would already have to be a general account of what meaning is — there would have to be a theory of meaning. Unfortunately the possibility of our establishing such a theory seems almost as remote now as it ever was. Some go so far as to claim that 'meaning' is a discredited notion,⁸ while others, who seek to explain it, differ radically in their proposals.⁹ What is frequently overlooked, however, is that simply refusing to talk about meaning does not constitute an answer to the challenges of Feyerabend, Kuhn, and those who share their views, for in doing so we give up the attempt to decide which terms from different theories mean the same. Any scepticism about meanings must also tell against establishing commensurability.

Later in the thesis I shall have more to say about meaning. At

present though we must keep in mind the question of what we want a theory of meaning for. In this thesis I have been, and will be, urging that a realist view of the cognitive status of scientific theories is not only intuitively appealing but also suggests a plausible account of the growth of scientific knowledge. According to the realist, the scientist aims at finding better theories — ones that lead to greater predictive success, are simpler, and so on — about the kinds of things there are in the world, and science progresses because this methodology works. The aim of the scientist is to find a true theory or description of the regularities of the world, and even though it might later be denied that there is what a theory says there is, at the time the theory is propounded it is at least claimed that it truly describes the real world. The initial problem is to give enough semantic theory to get the realist view going. Let me explain this further.

The realist's primary interest, as was noted at the end of the previous chapter, is not in the concept of meaning but in the concept of reference. His account, as we shall see, explicitly mentions the extensions that scientific predicates have. The core of the realist answer to problems concerning the growth of science is that in important cases of theory change the theories involved are still about the same things. By this is meant not just that they involve more or less the same sorts of things, in the sense in which two theories about acids, say, are both theories about chemical compounds, but that in many cases the properties are the same and even the most theory-laden of terms still pick out the same objects. Thus, Bohr's theory of the hydrogen atom mentions electrons — so does current sub-atomic physics. Newtonian physics talks about length (amongst other things) — so does

Einsteinian physics. The American geneticist H.J. Muller postulated theories about genes in the 1920's and 1930's — so do modern molecular biologists. It does not just happen that the same terms were taken over by the later theories — there is a definite sense in which they use the terms to talk about the same things. Or so the realist claims. We shall have to be a good deal more explicit about what this sense is, but let us first look at some historical background.

There is a firm basis for the distinction between meaning and reference in the philosophy of language. It was first explicitly discussed in Frege's justly celebrated paper "Über Sinn und Bedeutung," published in 1892,¹⁰ although he had mentioned it in a paper of the preceding year.¹¹ He presented the distinction as a means of resolving certain problems associated with proper names. What he wanted to explain was how it is that two names which refer to the same object may yet be used in an identity statement to convey factual information, whereas the use of the same name twice does not. That is, he wanted to explain why "a=a and a=b are obviously statements of differing cognitive value."¹² His answer relies on distinguishing between the "Sinn" or "sense" of a proper name and its "Bedeutung" or "reference". The "reference" is that which we use the name to talk about. Its "sense" is rather more complex and is the means by which the reference of a name can be determined. Understanding the sense of a name is what is required in order for us to decide which object it refers to. As an illustration, consider the two names 'Hesperus' and 'Phosphorus'. At one time these were used to refer to what were thought to be distinct stars in the sky. It was later discovered that they were both names of the same object, the planet Venus. The identity statements 'Hesperus=Hesperus' and

'Hesperus=Phosphorus' are both true but clearly differ in cognitive value — after all, it was an astronomical discovery that led people to say the latter was true. On Frege's account, the difference in cognitive value results from the fact that, whilst having the same reference, the two names differ in sense. In order to be able to decide what each of the names refers to, though, we have to grasp their senses.

It should not be thought that sense and reference are in some way constitutive of meaning. For Frege, reference signifies the relation between a word and an object and so forms no part of what we might ordinarily think of as meaning. Sense is more closely related to meaning. Knowing the sense of a word — understanding it — enables one to decide which object, if any, the word applies to; but there being a relation in the first place is something different. Failure to appreciate this point amounts to failure to understand Frege's solution to his original problem. Frege does not actually talk about sense being part of a more general notion. Nevertheless, it is clear that within the intuitive notion of meaning he thought that a distinction could be drawn between sense and two other ingredients — "force", as in the difference between asserting something and asking whether it is true, and "tone", as in the difference between 'sweat' and 'perspiration'. As Frege remarks, however, the latter ingredients concern "poetic eloquence ... and must be evoked by each hearer or reader according to the hints of the poet or the speaker."¹³ Presumably poetic eloquence will not be reflected within the general area of discourse with which we shall be concerned, namely that of scientific theories. Our investigation of meaning, then, will rather be an investigation of sense, though we shall treat topics which Frege never considered and say things with which he might well have disagreed.

Before suggesting how we might extend Frege's original insight, let me say a few words about terminology. Many scientific terms are predicates, i.e., expressions of the form '... is a ψ ' or '... is ϕ ', where the blank is to be filled by a singular term such as a name in order to yield a complete sentence. The true sentence 'Hesperus is a planet', for example, is made up of the singular term 'Hesperus' and the predicate '... is a planet'. Other examples of singular terms are demonstratives like 'this cloud chamber' and 'that pendulum', and definite descriptions like 'the 20cm.reflector telescope in the St. Andrews University observatory'. In talking of a singular term I shall sometimes say that it denotes an object, by which I mean that it refers to the object, and sometimes that such an object is the referent of the term. Other terms used so far which function as predicates are 'gene', 'electron', and 'acid'. I shall also include events, like 'an evolution of gas', and instances of matter, like 'the electroscope leaves', as particulars over which predicates range. In this way 'chlorine', 'gold', etc., can be counted as predicates. The notion of reference is to be understood broadly as signifying not only the relation between a singular term and an object but also the relation between a predicate and a set of objects, this set being the extension of the predicate. So when I talk about a theory of reference for scientific terms this is sometimes to be understood as meaning a theory of the extensions of scientific predicates.

Returning to the problem of theory comparability, suppose that theory T^1 has the following empirical consequence:

$$(i) \quad (x)(Px \supset Qx)$$

Suppose also that an apparent rival, T^2 , has the empirical consequence:

$$(ii) \quad (\exists x)(Px \ \& \ \sim Qx)$$

Ordinarily we should maintain that, since (i) and (ii) are logical contradictories, T^1 and T^2 are in conflict. In maintaining this we are committed to holding not that both 'P' and 'Q' have the same sense in T^1 and T^2 but that they have (more or less¹⁴) the same reference. The reason is that, in extensional logic, to say that (i) and (ii) are contradictories is to say that they could not both be true under any uniform interpretation of the predicates, where by 'uniform interpretation of the predicates' is meant: given a non-empty domain of objects D , and an interpretation I which assigns sets of objects to predicates as their extensions, all occurrences of the same predicate letter under an interpretation I are to be construed as having the same extension. If it can be established that both 'P' and 'Q' have the same extension as they occur in T^1 and T^2 , then they can be represented, as in (i) and (ii), by the same predicate letter in logical form. Finally, no uniform interpretation of the predicates could possibly make both (i) and (ii) true.

Apparently, then, the realist's main task is to establish what the extensions of predicates from different scientific theories are and to express this in such a way as to permit comparison. The relativist, or super-realist, denies that such inter-theoretic comparison is possible. We have seen certain problems that arise for the relativist, but it does not thereby follow that realism is vindicated. As I see it, the relativist's denial is a challenge to the realist to justify his claim that there is a fact of the matter as to whether competing or successive theories are about the same things.

Section (ii): Four Questions for Realism

Let me now try to spell out realism in more detail. As I said before, my main aim is to explicate the realist account of the development of science, and the central thesis of this account is that there are many cases where competing or successive theories are about the same things. Evidently then a realist will have to be able to answer the question: how can we compare the extensions of relevant predicates from different scientific theories? What we have to have a better idea of though is which predicates are the relevant ones. If we look at some of the typical problematic cases of theory change, we come across questions like 'Are the formative elements postulated by Mendel the genes studied by modern molecular biologists?', 'Were the atomic theories of Dalton and Avogadro about the same things?', and 'When Newton used the term 'gravitational field', did it apply to anything recognized in Relativity physics?'. Less problematic cases, i.e., ones where sameness of extension is more obvious, are Ptolemaic and Copernican theories of the planets, Aristotle's theory of the brain and that of contemporary neuro-physiologists, and even the discovery that whales are mammals and not fish. In all of these cases, the relevant predicate — 'gene', 'atom', 'gravitational field', 'planet', 'brain', and 'whale' — is a natural kind predicate. Such predicates, which can be correctly applied to objects on the basis of their physical properties, occupy a central place in scientific theorizing. At the beginning of Chapter 5, I shall talk about them in more detail. I shall also mention examples of theory change, to be discussed in the final chapter, which involve predicates not associated with natural kinds.

For the present though I wish to restrict my investigations to the central cases of natural kind predicates.

The most general question for a realist account of the growth of scientific knowledge then is:

- (1) How can we compare the extensions of relevant natural kind predicates from different scientific theories?

This question can be taken in a formal sense or in a material one.

In the formal sense it is a question about the logic of theory comparison — what language or theory should we use for comparing the extensions of natural kind predicates from two distinct theories?

The next section will contain my answer to this formal question.

In the material sense, (1) is a question about the means that are to be employed in actually ascertaining what the relevant predicates of different theories have as their extensions. In concentrating on reference, the realist has to be able to meet the challenge posed by a particular form of the incommensurability thesis, viz., that there is no fact of the matter as to whether or not predicates from different scientific theories have the same extensions. How is he to meet this? He has to give a general account of how the extensions of relevant predicates can be discovered, i.e., an account of how we are to decide which things a particular predicate can be, or was, correctly applied to.

In pursuing an investigation into the epistemological question of what a particular scientific predicate has, or had, as its extension, the realist will sometimes be led to consider terms like 'whale', 'planet' and 'brain' which are used outside of science. Many natural kind predicates are like this. Often this feature will be of some use

in understanding what the extensions of terms used in past scientific theories were. The more widely a term is used within a linguistic community, the more information one who wishes to understand what it means and what it refers to will have available. A study of the use of a natural kind predicate, then, is best construed as a study of its use not just within a particular scientific theory, but within a linguistic community. I shall discuss use and understanding more fully in subsequent chapters. For the present, we are now in a position to formulate the second of the four questions underlying the realist's account of theory change:

- (2) How can we discover which objects belong to the extension of a natural kind predicate as that predicate is used within a linguistic community?

A complete answer to question (1) depends on there being an answer to question (2), as well as on the formal problem of deciding on a language for theory comparison. Only if answers to both of these are achieved will the realist be able to meet the challenge posed by the thesis of incommensurability.

What sorts of considerations will have to be taken into account when it comes to answering (2)? I have discussed the Fregean distinction between the sense of a term and its reference or extension, and noted that associated with the sense is a criterion for determining the reference or extension. I also said that sense could be considered as the primary notion constitutive of meaning. Thus we might expect that the sense of a scientific predicate will be given, at least in part, by the theoretical principles used in stating the theory, or theories, in which it occurs. But the realist has to be extremely careful here.

As we have seen, there are problems involved in making precise the notions of 'partial specification' or 'partial interpretation'. More fundamental though is the following difficulty. Suppose for the moment that the realist wishes to construe the statement that the sense of a scientific predicate is given in part by theoretical principles as saying that there is, associated with the sense, a criterion, satisfaction of which is a necessary condition for an object's belonging to the extension. Then in nearly every case of theory change he will not be able to show that successive theories are about the same things! Putnam has offered a vivid example:

Bohr assumed in 1911 that there are (at every time) numbers p and q such that the (one dimensional) position of a particle is q and the (one dimensional) momentum is p ; if this was part of the meaning of 'particle' for Bohr, and in addition, 'part of the meaning' means 'necessary condition for membership in the extension of the term', then electrons are not particles in Bohr's sense, and, indeed, there are no particles 'in Bohr's sense'. (And no 'electrons' in Bohr's sense of 'electron', etc.) None of the terms in Bohr's 1911 theory referred! It follows on this account that we cannot say that present electron theory is a better theory of the same particles that Bohr was referring to.¹⁵

From this the realist does not conclude that the terms of Bohr's theory were like, e.g., the term 'phlogiston', which we now say cannot be correctly applied to anything at all. He prefers to say that some of Bohr's beliefs about electrons, etc., and hence some of his theoretical postulates about them, were mistaken. Similar remarks might be made concerning, say, Ptolemy's beliefs about the planets — he thought they orbited the earth — and Muller's beliefs about genes — he thought they were composed of proteins. But to maintain this, whilst at the same time holding that modern researchers are talking about the same

things, means that satisfying the full sense of a term cannot be thought necessary for an object's being a member of that term's extension, or, in the case of a singular term, for its being referred to by that term.

Two points follow from this. The first is that we shall not be concerned with what Muller, Bohr, Ptolemy, etc., "had in mind" or "intended to refer to" when they used terms like 'gene', 'electron' and 'planet'. What they had in mind may be revealed by more general considerations to do with what they said, how they acted, and what they believed. Our concern, on the other hand, is with what they succeeded in referring to, where success is judged relative to our present understanding. The most striking cases of this are ones like theories claiming to be about caloric, phlogiston, magnetic flux, and the luminiferous ether, where we now say that these terms fail to refer or have an extension.

The second point is that it is incumbent upon us, in so far as we want to continue to use the notion of 'sense', to reject the view that there is always a single core of descriptions associated with a term, satisfaction of which is a necessary and sufficient condition for an object's being a member of that term's extension, or for its being denoted by that term. We have to allow for error, for mistaken belief, not only on the part of our predecessors but also on the part of ourselves. Putnam has attempted to avoid this difficulty by abandoning the notion of 'sense' altogether and using what has come to be known as a causal account of reference.¹⁶ In Chapter 5, I shall criticize this strategy and offer an alternative based on a cluster theory of reference.

The causal account emphasises features of our use of proper names and natural kind predicates which are relevant to answering the epistemological question of how we can discover what a term refers to or has as its extension. If we can discover certain facts about the causal origin of a theory — what apparatus and data a scientist or group of scientists had available prior to advancing a theory — this might enable us to rule out some conjectures as to a predicate's extension, or enable us to settle more quickly on others. In this way the causal account, so I shall argue, is relevant to answering question (2). What it is not relevant to though is answering the conceptual question of what it is for a proper name to refer or for a predicate to have an extension. Clearly this latter question has to be answered before the former can be; we have to know what it is for a predicate to have an extension before we can decide what the extension of a particular predicate is. Hence we have a third question for realism:

(3) What conditions have to be satisfied by a natural kind predicate

' ϕ ' and an object a in order for ' ϕ ', as it is used within a linguistic community C , to be correctly applied to a ?

It is important to make a clear distinction between questions (2) and (3). (2) is an epistemological question, (3) a question of conceptual analysis. By way of analogy, consider the two questions 'How can we discover who the father of a given person is?' and 'What is it to be a father?'. The former is epistemological, the latter conceptual. There are various tests and criteria for discovering paternity, but obviously their very existence presupposes that it is understood what it is to be a father.

Although there is an important distinction here, we might expect

that once we have formulated an answer to the conceptual question this will suggest criteria for answering the epistemological one. An explanation of what it is to be a father clarifies the relation of paternity and so enables us to devise criteria for discovering who a person's father is. In Chapter 5, I shall argue that a predicate ' ϕ ' can be correctly applied to an object if and only if that object satisfies a suitable majority of those descriptions believed to be true of ϕ 's. This suggests that, in order to discover what a particular natural kind predicate has as its extension, we shall have to determine what is or was believed, by users of the predicate, to be true of things of that kind. But this is not to say that 'belief' is, as it were, the only category with which the realist has to work. For it is at this stage, i.e., the stage of answering a particular instance of (2), that he can make good use of the causal account of reference. In Chapter 6, I shall also suggest other criteria that might aid him in discovering the extensions of scientific predicates.

Questions of sameness or difference of extension relate primarily to understanding previous scientific theories. The problems of understanding the language associated with a theory will also be compounded in these cases. The realist might even be forced to recognize that, despite the constraints alluded to above, he cannot always determinately translate from another language to his own. He would then have to decide if such indeterminacy affects the question of what reference or extension a term has, or whether it merely affects the sense. I shall have a lot more to say about these issues later, but first we need to pose a final question.

Returning to the level of question (3), let us consider a particular predicate ' \underline{P} ' which is part of some language L_1 used by

members of a given linguistic community. To say that 'P' can be correctly applied to some object a₁ is to say that 'Pa₁' is a true sentence of L₁. This concept of truth will be discussed at length in Chapter 4. I introduce it now so as to give some idea of the close relation between the theory of reference and the theory of truth. This relation suggests that we might paraphrase question (3) as 'What conditions have to be satisfied by a natural kind predicate 'φ' of language L and an object a in order for 'φa' to be a true sentence of L?'. In this way we see that in coming to understand what it is for a predicate to have an extension we come to understand what it is for a statement to be true.

A central tenet of the realist's position is a correspondence theory of truth, i.e., a theory whereby what it is for a statement to be true is explained in terms of a relation, a correspondence relation, between that statement and something else. In Chapter 4, it will be shown that such a theory can be turned into a recursive definition of truth. But this raises the question of what it is for there to be such a recursive definition, what makes the definition possible? In discussing these problems I shall focus my attention on the question:

- (4) What conditions have to be met in order for truth to be recursively defined?

My answer to this most fundamental of the four questions will be that, where we are concerned with some natural language, truth can only be recursively defined if there is a determinate relation of reference between singular terms of the language and objects in the world, and between predicates of the language and sets of objects in the world. Reference, I shall say, in this way underwrites truth. Here we have

reached the foundation of realism. An explanation of the concept of truth, of what it is for a singular term to denote or for a predicate to have an extension, of how we can discover which objects a given predicate can be correctly applied to, and hence whether or not predicates from different scientific theories have the same extensions, all depends on there being a primitive relation of reference between language and the world.

That there is such a determinate relation might seem obvious. Certainly we have seen no cause to doubt that there is so far; even the relativist or super-realist positions of Feyerabend and Kuhn assume that there is. The realist represents the scientist as one who postulates theories describing the real world. Moreover, the scientist is not usually thought to be privileged in thus making contact with "external reality" — most of us, most of the time, think, act and talk as though we do. But how determinate is the relation between language and the world? In the next chapter, I shall consider a forceful argument presented by Quine for the view that even though there is some relation between language and the world it is not that which philosophers since Frege have called reference. In a sense I shall be dealing with a fifth question for realism. If we find grounds for doubting that there is a determinate relation, then the other questions lapse.

Having posed four questions and sketched their interrelations, let me now outline the structure of the remainder of the thesis. To give some indication of the strategy we need to adopt, the next section contains an answer to the question of the "mechanics" of theory comparison, of the language that is to be used. My remarks will be based on consideration of an actual example from the history of the

atomic theory. In the following chapter, I shall discuss Quine's argument that reference is inscrutable. Replying to this, I shall attempt to show that it is not so, either in our own home language or when it comes to translating an alien language. Starting in Chapter 4, I shall work systematically through the four questions in reverse order, beginning with the relations between truth and reference. Chapter 5, as I said before, will contain a cluster theory of reference for natural kind predicates. Before this is proposed, however, I shall draw a strong analogy between natural kind predicates and proper names, and defend a cluster theory of reference for proper names. Much of the support for the natural kind predicate theory will derive from arguments in defence of the proper name theory. The penultimate chapter will investigate more closely how we can understand what earlier scientific theories were about. I shall also look at how this task is affected by Quine's arguments for the "indeterminacy of translation of sentences." Finally, in Chapter 7, I shall discuss some more examples and consider the contrast between saying that earlier scientists had some incorrect beliefs about things we now countenance, and saying that they failed to refer to anything at all. This should give us a firm grasp of what is required to answer question (1).

The first person to explicitly suggest using the theory of reference in an attempt to rebut incommensurability was I. Scheffler, in his book Science and Subjectivity, which appeared in 1967. To the best of my knowledge, however, neither Scheffler nor anyone else has drawn attention to, let alone fully discussed, the set of points just made. They are, nevertheless, crucial to a thorough examination of the foundations of realism in the philosophy of science. Important

arguments which bear on some of these points have been raised by Quine, Evans, Putnam, Field, and others — we shall consider them in due course — but nowhere do they give a unified account of the sort I shall be presenting here. The more familiar pattern of discussion seems to have been set by Scheffler himself, who argues at length in his book that sameness of reference is what is required for comparing theories, yet leaves it totally unclear how the reference of an expression is to be discovered! Recently an attempt has been made by K. Parsons to fill this lacuna and hence to answer questions like (1) and (2).¹⁷ I shall consider it closely in the remainder of this chapter. Parsons goes beyond other followers of Scheffler in that she recognizes that discovering what the extension is of a scientific predicate as used in a theory requires us to say something about how the predicate comes to be understood. However, rather than addressing herself directly to question (3) and the conceptual problem of what it is for an object to belong to the extension of a predicate, she only offers some vague suggestions for deciding the epistemological question of how to discover what a particular predicate's extension is. The inadequacy of these suggestions will be made apparent when we look at her reply to one argument for incommensurability.

Section (iii): Saying What the Extension of a Predicate is

To get a better feel for the problems confronting a realist account of the growth of scientific knowledge, let us turn immediately to question (1) of the last section. Suppose we have two scientific theories T^1 and T^2 , and we are interested in comparing the extensions of predicate 'P' — a predicate used in T^1 — and predicate 'Q' — a

predicate used in T^2 . Sometimes 'P' and 'Q' will be the same term, as with the theories about genes and electrons mentioned above. This is not so for the example I shall be concentrating on in this section. In cases where T^1 and T^2 are rival or successive theories, the realist will aim to establish that whatever is a P according to T^1 is a Q according to T^2 . But how should we express this in logical form? At first it seems correct to say that in such cases he wants to establish the truth of the proposition

$$(x)(Px \equiv Qx)$$

When we think about it, though, this proposition must strike us as an unholy mixture. Since 'P' and 'Q' are terms belonging to different theories, how can they be used in the same proposition, bound by the same quantifier? The crucial term in this question is 'used'. To see why, let us descend from the realm of logical form and make a case study. To begin with, I trust that I shall be forgiven if I beg some questions and make some assumptions about theory comparability in order to describe the case.

In 1808, John Dalton published Part I of his New System of Chemical Philosophy in which he propounded a theory of chemistry based on a molecular theory of matter. The postulates which formed the basis of his molecular theory were essentially satisfactory and correct. Unfortunately, however, they were not in themselves sufficient to make possible a calculation of atomic weights, which was necessary in order to show in what way atoms of different gases were different. To fill this lacuna, Dalton boldly assumed an arbitrary set of maxims which amounted to a rule of greatest simplicity with respect to the ratios in which elements combined to form compound molecules,

If there are two bodies, A and B, which are disposed to combine, the

following is the order in which the combinations may take place, beginning with the most simple, namely:

1 atom of A + 1 atom of B = 1 atom of C, binary.

1 atom of A + 2 atoms of B = 1 atom of D, ternary. ...

The following general rules may be adopted as guides in all our investigations respecting chemical synthesis.

1st. When only one combination of two bodies can be obtained, it must be presumed to be a binary one, unless some cause appear to the contrary. ...¹⁸

The trouble with these maxims is that they brought Dalton's theory into conflict with some carefully determined experimental results of the French chemist Gay-Lussac. These were encapsulated in his Law of Combining Volumes: when gases react chemically, the proportions by volume measured at the same temperature and pressure bear simple whole number relations to each other, and the volumes of the products (if gaseous) measured under the same conditions also bear simple whole number relations to the volumes of the reacting gases. When oxygen and hydrogen react to form water, for example, one volume of oxygen and two volumes of hydrogen yield two volumes of water vapour.

If Dalton had accepted that there is some simple relation between volumes of different gases and the number of "atoms" they contain, he might have been able to explain these results. But he had at least three reasons for not doing so. In the first place he thought that "atoms" of different elements have different sizes, and so believed that different numbers of them would occupy the same volume. Secondly, there would have been cases where, in order to reconcile his maxims with Gay-Lussac's law, he would have to have accepted that certain of what he regarded as atoms — oxygen molecules, for example — were composed of two like parts. From the beginning, though, Dalton had been impressed by Newton's work on electricity and magnetism, and he adopted Newton's

model of air as an elastic fluid constituted of like atoms which repel each other by a force increasing as their distance diminishes. 'Like repels like' was, for Dalton, a universal truth. Finally, Dalton claimed to have experimental evidence of his own that conflicted with Gay-Lussac's law. Thus, Dalton remained steadfastly opposed to the law.¹⁹

In 1811, Amadeo Avogadro, an Italian physicist, published an obscure paper entitled "Essay on a Manner of Determining the Relative Masses of the Elementary Molecules of Bodies, and the Proportions in Which They Enter into These Compounds."²⁰ He pointed out that to deny that the number of molecules contained in a given volume at a given temperature and pressure were the same for different gases would have made it virtually impossible to account for the observed regularities of reactions. Avogadro went on to propound a hypothesis, now referred to as his law, that equal volumes of all gases, under the same conditions, contain equal numbers of molecules. He added, as a corollary to this, what we should now express by saying that the molecules of elementary gases were usually composed of more than one atom. On the basis of the hypothesis it was possible not only to explain Gay-Lussac's results but also to provide a method of finding true molecular formulae without arbitrary maxims of simplicity. The advances made in the chemistry of the late nineteenth century were the result.

How does all this illustrate our problem? Well, we have two theories about the molecular composition of gases — Dalton's and Avogadro's. Let us denote the former as T^1 and the latter as T^2 . From the point of view of our own science, it seems natural to say that one salient difference between T^1 and T^2 is that T^2 distinguishes atoms from molecules whereas T^1 does not. Of course, it is not quite as simple as this. Although Dalton uses 'atom' and 'molecule' more or less inter-

changeably, he also uses the terms 'ultimate particle' and 'simple elementary particle' to denote atoms or molecules of compounds.²¹

Avogadro also employs a wide range of terminology, distinguishing atoms of elements from molecules of elements and molecules of compounds.²²

And this is only how we would characterize their vocabulary if we had some way of comparing terms from different theories.

For the sake of simplicity, I shall concentrate on Dalton's term 'atom', which he uses in presenting his maxims of simplicity, and Avogadro's term 'molecule', which occurs in his hypothesis. Were they, in using these terms, talking about the same things or not? How are we to decide? It seems plain that when Dalton uses the term 'atom' to talk about the elementary particles of gases, he is referring to what Avogadro and modern chemists would term 'molecules'. Because of this it seems plain also that the claim Dalton would make using the statement 'No atoms of gases are composed of two like particles' is directly contradicted by the claim Avogadro and modern chemists would make using the statement 'Some molecules of gases are composed of two like particles'. And it is this which inclines us to say that Dalton's theory was factually incorrect, i.e., that it led to false statements about how the world is.

The realist's strategy here is to try to establish that 'atom of a gas' for Dalton had the same extension as 'molecule of a gas' for Avogadro. Substituting the letters 'P' and 'Q' respectively for these predicates, we may say that what he tries to establish is the truth of the statement

(1) 'P' and 'Q' are co-extensional.

If this can be established, then the two claims may be written in logical form using the language of T^2 — the language in which 'Q' is used — as

$$(2) (x)(Qx \supset \sim Dx)$$

$$(3) (\exists x)(Qx \& Dx)$$

where 'D' is the predicate of T^2 'is composed of two like particles', and where it is assumed that there are some things of which 'Q' is true. From this the realist can conclude that the two theories were rivals in that they offered contradictory accounts of the same objects — what Avogadro's theory identified as molecules of gases. Since we now accept most of the things Avogadro said, this provides the basis for a criticism of Dalton's theory; (3) is true and hence (2) is false, Dalton's maxims also were mistaken. So some of Dalton's beliefs about molecules were incorrect.

How is the realist to go about establishing the truth of (1)? There are two problems at issue here, one corresponding to what I called the formal sense of question (1) from the previous section, and one corresponding to what I called its material sense. For the time being I shall concentrate on the first of these problems.

Evidently the realist wants to be able to assert both "The extension of 'P' is ..." and "The extension of 'Q' is ..." in the same language, since then the quantifiers will range over the same ontology, thus permitting comparison. It would seem that the obvious way to do this is to construct a metalanguage in which one can make statements about the referents of the singular terms and the extensions of the predicates of an object language. This concept of formal semantics was first articulated by Tarski.²³ As regards its application in the logic of theory comparison, Putnam has remarked, "if one has a scientific language L containing the term 'electron', then one can certainly construct a metalanguage ML over it à la Tarski, and define 'reference' in such a way that "'electron' refers to electrons" is a trivial theorem."²⁴

While this provides a way of saying what the extension of a predicate is, it does not immediately solve the problem of comparing the extensions of predicates from different scientific theories. If we have two such theories, T' and T'' , we cannot assume that they are to be associated with the same formal language. If we associate them with different languages L' and L'' respectively, though, they will also be associated with different metalanguages ML' and ML'' . Using Putnam's example, we can say in ML' "'electron' refers to electrons", intending that 'electron' in the sense of T' refers to what are identified as electrons in the theory T' , and in ML'' we can say "'electron' refers to electrons" intending that 'electron' in the sense of T'' refers to what are identified as electrons in the theory T'' . But as Putnam concludes, "there is no ML in which we can even express the statement that 'electron' refers to the same entities in T' and T'' ." ²⁵

The realist will need to go a little further here and justify the use of the same metalanguage for talking about both theories. Following this line of argument, Parsons has suggested an answer which relies on a general criterion of what it is for a predicate to be true of an object. She expresses this criterion as,

for any predicate expression ' P ':

For any a , ' P ' is true of a in L if and only if a is an F (for some F) ²⁶

By way of example she suggests, "for any individual a , 'lithologist' is true of a in English if and only if a studies rocks." ²⁷ These lead her to present the following statements for comparing predicates,

(4) $(x)(\text{'P' is true of } x \text{ in } T^1 \equiv Fx),$

(5) $(x)(\text{'Q' is true of } x \text{ in } T^2 \equiv Fx),$

from which she infers

(6) $(x)(\text{'P' is true of } x \text{ in } T^1 \equiv \text{'Q' is true of } x \text{ in } T^2). ²⁸$

Understanding T^1 as Dalton's theory, T^2 as Avogadro's, and 'P' and 'Q' as before, these three statements would seem to be just what the realist is after, for (6) is then simply another way of capturing the import of (1). Both (4) and (5) are statements belonging to the same metalanguage and so the inference is justified. The remaining problem is then whether there is a suitable predicate to substitute for 'F'.

What is important about Parsons' suggestion and her criterion is that they show how useful it can be to conceive of truth as a three-place relation holding between a predicate, a set of objects and a language. In the previous section I mentioned the conception of truth as a two-place relation holding between a statement and a language. Seeing it as a three-place relation is an obvious extension. This, I think, makes it clear how the realist should answer question (1) when understood in the formal mode. In order to resolve problems of co-extensionality between predicates of different theories, we need to be able to specify their extensions using the same metalanguage. In this way the variables all range over the same ontology. The question which remains is (1) understood in the material mode. How exactly can we establish the truth of the likes of (4) and (5)?

One of the main difficulties is to decide on a predicate 'F'. I have already pointed out that, where we now think a theory is mistaken on some points, we cannot in general hold that a necessary condition for an object's belonging to the extension of a predicate of the theory is that it fully satisfy the sense of the predicate as specified within the theory. In the present case, for example, the realist's view of how science progresses could not be made out if part of the sense of 'F' were assumed to be 'is not divisible by chemical means', for although Dalton appears to have believed this to be true of what he called "atoms",

it is not true of what Avogadro and present-day chemists call "molecules". It would be more plausible if 'F' were interpreted as 'is a smallest particle of a gas to have all the chemical properties of the gas', but rather than compare such specific alternatives, the question I want to press is how we could know any one of them to be correct. No doubt there are cases of theory change where no suitable predicate can be fixed on; after all, the realist is only claiming that his account is one of what typically happens. But what conditions have to be fulfilled in order for a predicate to be suitable, and how can we, in practice, interpret one theory using the language of another?

Parsons, in her discussion of the Dalton/Avogadro case, does not address herself directly to these questions, although she does make several points which relate to them. Her basic approach to problems of change of extension is to advance empirical "meaning hypotheses" concerning the extensions of the relevant predicates.²⁹ These hypotheses are then tested against the "informational backgrounds" of the theories concerned. If judged correct as a result of these tests they are defended, where necessary, against the actual arguments given by the adherents of the theories for contrary assignments of extensions.³⁰ Expressing the last step in the terms I have used in this chapter, what it amounts to is deciding which of the beliefs the adherents of the theories had about, say, molecules were false.

The idea of advancing empirical hypotheses about extensions of predicates seems to me to be a good one. As I have already emphasised, the realist's case should be based on contingent facts; specifically, the truth of his claims about the growth of science depend on what information we now have available about past scientific theories. He must, therefore, be prepared to revise an assignment of extension to a

predicate of a past theory in the light of further evidence. There is also something to be said for a notion like that of an informational background. In Chapter 6, I shall talk about it more. It becomes particularly important when we ask questions like how Dalton was able to explain, in his own day, what he was talking about when he used a term, such as 'atom', to which he gave a particular sense within his theory. How did other scientists come to understand Dalton? In particular, how could Avogadro have come to understand him? The answer, I shall argue, is that there are certain conditions that have to be satisfied in order for a predicate or singular term to be successfully introduced into, and used within, a scientific theory. Here I shall appeal to a causal account of reference. We also have to remember that Dalton's theory did not suddenly appear, phrased in its own scientific language, complete unto itself. Most of the terms he used had been used in other theories before. It is rather that there already was a scientific language which was linked to a wider natural language and which Dalton augmented by introducing terms with particular intended senses.

This is getting away from the substance of Parsons' article though. There the only characterization given of the notion of an informational background is that it "contains the best substantiated theories of the day."³¹ The term 'best substantiated' is most unclear, but leaving this aside difficulties seem to arise in cases of theory conflict when neither theory is obviously better substantiated than the other. According to Parsons' characterization, neither theory would then appear in the informational background, so in principle there would be no way of ascertaining whether such rival theories are about the same things. And this is one of the main problems Parsons sets out to solve!

What is more, the very example of the Dalton and Avogadro theories would seem to be a case in point, for in the fifty years between 1811 and 1861, both were known to many chemists, yet neither enjoyed more than limited acceptance.

A more general criticism of Parsons' article is that she takes too much for granted. For example, she says,

In practice, a meaning hypothesis will be explicitly made only for a fairly small number of key terms — those about which there is some question — while others are assumed to have the same meaning in meta- and object language. This assumption, of course, constitutes an implicit meaning hypothesis.³²

What Parsons says might be quite true of what will happen in practice using her approach. But what is it that justifies the implicit hypothesis? What entitles us to assume that the extensions of the many predicates for which no hypothesis is made remain the same? It seems that what is lacking is a theory of interpretation, a theory of how we can interpret one scientific theory, like Dalton's, using the language of another, like that used to talk about Avogadro's. Finally, there is also the presupposition that there is a well-defined relation between predicates of scientific theories and objects in the world. Parsons makes use of this in proposing her general criterion of what it is for a predicate to be true of an object. As I said, I think this criterion is both correct and useful in the logic of theory comparison. I argued in section (ii), however, that a thorough justification for realism has to refute any claims that there is no such determinate relation to be made the subject of a criterion. In the next chapter I shall look in detail at one such claim.

CHAPTER 3 THE 'SCRUTABILITY' OF REFERENCE

Section (1): Quine on Translation and Meaning

We use a metalanguage in order to specify the relations between singular terms of an object language and their referents and between predicates of an object language and their extensions. The metalanguage enables us to mention the expressions of the object language. Those cases in which we are particularly interested are where the object language contains a scientific theory, i.e., all of the terms used in stating the theory are assumed to belong to the language. When we wish to talk about the terms of the theory, any metalanguage constructed over the object language in the Tarskian manner will suffice. Often, however, we wish to know what relations hold between expressions of one theory and those of another; relations such as coreferentiality and coextensionality. This other theory might be a more comprehensive theory in the sense that its domain includes that of the first. On the other hand, it might just be another theory related more loosely; the domains overlap. Another alternative still would be that there is no overlap whatsoever, in which case the theories have different domains. In order to permit comparison, though, we must in all cases be able to talk about the terms of both theories using the same metalanguage. Otherwise we shall be unable to say anything at all about the relations that interest us, not even, e.g., that the extensions of certain predicates are different.

The idea that we must use a background language for talking about connections between words and the world has suggested to some people — notably Quine — that the only relations which really make

sense are those between one language and another, not those between language and the world. Perhaps the key points in his argument for this view are that, from the standpoint of a different background language, the connections might be different, and that we have no telling reason for preferring one background language to another. Using a favourite example, Quine puts the points like this,

When we ask, "Does 'rabbit' really refer to rabbits?" someone can counter with the question: "Refer to rabbits in what sense of 'rabbits'?" thus launching a regress; and we need the background language to regress into. The background language gives the query sense, if only relative sense; sense relative in turn to it, this background language.¹

From this he draws a radical conclusion, "what makes sense is to say not what the objects of a theory are, absolutely speaking, but how one theory of objects is interpretable or reinterpretable in another."²

Quine is urging here that severe restrictions apply to what can be said about relations between words and the world. There is no fact of the matter about reference, only "relative facts", facts relative to a background language. This might even be construed as a radical form of the incommensurability thesis, for it questions the whole notion of reference, of whether or not there is such a determinate relation between language and the world. Quine's argument, then, is directed at the very foundations of realism as I have explained it.

Quine has argued at length and often that language is indeterminate in various ways. As he has recently emphasised, however, there are two main strands of argument for this conclusion which need to be distinguished.³ The first turns on the underdetermination of physical theory by all possible evidence. According to Quine, "theory can still vary though all possible observations be fixed. Physical

theories can be at odds with each other and yet compatible with all possible data even in the broadest sense. In a word, they can be logically incompatible and empirically equivalent."⁴ From this point, on which Quine expects wide agreement, he claims to infer that our translation of a foreigner's theory will permit at least as much variation, in that it will be indeterminate which theory should be attributed to him given a determinate translation of his observation sentences. When a 'foreigner' is taken to be an earlier theorist, this argument obviously has considerable relevance to our attempt to understand previous scientific theories. I shall consider it in detail in Chapter 6. In reply it will be maintained that, given the underdetermination of theory by all possible evidence, if this results in the translation of previous scientific theories not being determinate, such lack of determinacy has not been shown by Quine to affect the extensions the predicates of those theories are said to have. In this chapter I shall confine myself to the second main strand of argument Quine adduces for indeterminacy of translation of sentences, that concerning what he calls "the inscrutability of reference of terms."

To begin with, Quine argues for the inscrutability of reference in the translation of predicates, or, as he usually calls them, general terms, from a foreign language for which we have no previously established dictionary. Thus, to take Quine's notorious example, suppose that a field linguist is wondering how to translate the foreign term 'gavagai' into English. He is able to establish inductively, beyond reasonable doubt, that a foreigner can be prompted to assent to the occasion sentence 'Gavagai' by the presence of a rabbit, or reasonable facsimile, and not otherwise.⁵ The obvious translation for the term would seem to be 'rabbit'. But, Quine notes, "a whole rabbit

is present when and only when an undetached part of a rabbit is present; also when and only when a temporal stage of a rabbit is present."⁶ There is no way of telling, simply by ostension, which of 'rabbit', 'undetached rabbit part' or 'rabbit stage' is the correct translation. Using Quine's terminology, 'Gavagai' in the foreign tongue can be established as having the same "stimulus meaning" as, i.e., as being "stimulus synonymous" with, the English 'Rabbit'. This means, roughly, that each occasion sentence would be assented to, or dissented from, by a speaker of the relevant language if asked under the same conditions of sensory stimulation. 'Rabbit', however, is in turn stimulus synonymous with the other English occasion sentences 'Undetached rabbit part' and 'Rabbit stage'. As Quine sees it, trying to settle behaviourally the extensions of the various terms constituting these behaviourally indistinguishable sentences is an insoluble, and therefore unreal, problem. Though it is obvious that the extensions of the various English terms must be distinct, there seems no aspect of actual or possible physical behaviour which tells us to which of them the linguist should map 'gavagai'.

How is the linguist to decide between the alternatives? Well, says Quine, he has to turn from ostension and observed behaviour to verbal stimuli and verbal behaviour. In English, our recognition of sortal terms — terms which divide their extensions — depends on grammatical particles and constructions like plural endings, pronouns, numerals, the 'is' of identity, and its related terms 'some' and 'other'. The linguist proceeds by abstracting what he takes to be similar particles and constructions from the foreign language and hypothesising how they are to be associated with the English ones. Such hypotheses of translation Quine calls "analytical hypotheses".⁷

The linguist may now ask the foreigners what he supposes amounts to the questions 'Same gavagai?' or 'How many gavagais?' Their answers would seem to settle matters. Quine agrees that to select a foreign word, say 'bleg', as the identity particle virtually fixes the translation; hence some analytical hypotheses are of crucial importance. But they are after all only hypotheses, themselves underdetermined by behaviour. Imagine that a foreigner sees, protruding from behind a small rock, a rabbit head and a tail. He is prompted to utter the sentence: 'Ip gavagai bleg op gavagai'. The linguist might equally well translate it as 'This rabbit is the same as that one' or as 'This undetached rabbit part belongs with that one'.

The possibility thus emerges that there are incompatible manuals of translation for predicates from a foreign language to our home language which would preserve invariant all the dispositions to physical and verbal behaviour on the part of speakers of the foreign language. The manuals are incompatible in the sense that they would carry a single predicate from the foreign language into various predicates of the home language which are not coextensional. So failure of reference, not merely failure of sense — which I shall later argue to be the import for sentence translation of the underdetermination of physical theory by evidence — is what is at issue.

Two features of this argument require further comment. The first is that Quine pays exclusive attention to the behavioural facts — ostension, verbal behaviour and the like — when considering how the linguist might fix on a particular translation manual. The study of language and the meanings it contains is, for Quine, primarily a study of behaviour, "language is a social art which we all acquire on the evidence solely of other people's overt behaviour under publicly

recognizable circumstances."⁸ Viewed correctly then, "knowledge, mind, and meaning are part of the same world that they have to do with, and ... are to be studied in the same empirical spirit that animates natural science. There is no place for a prior philosophy."⁹ Consequently there is nothing more to guide us in translating the language of another person than that person's dispositions to overt behaviour.¹⁰

The second feature is that Quine thinks it is obvious that there are no behavioural facts underlying the use of a term like 'gavagai' which could allow us to say that it is a predicate signifying one set under translation rather than another. That is, no behavioural facts give us a reason for preferring one translation to another. As we saw, appeal to the foreigner's dispositions to assent^{to} and dissent ~~to~~ from verbal stimuli fails to resolve the indeterminacy, since we then have to allow for the indeterminacy of translation of identity and other individuating apparatus. For someone to now say that there is some fact of the matter as regards the translation, even though it cannot be decided by appeal to behavioural facts, is for him to fall victim to what Quine derides as "the myth of a museum in which the exhibits are meanings and the words are labels."¹¹

Having noted these two features of the argument, let us see how Quine develops it. In place of the foreigner and translation from a remote language, let us put a neighbour of ours and our own home language. We become linguists ourselves, wondering how to 'translate' our neighbour's English discourse. Ordinarily we are guided by one compendious, analytical hypothesis, viz., the rule of homophonic translation: we equate expressions in our neighbour's mouth with the same strings of phonemes in our own. Sometimes there is a principle

of charity at work whereby we avoid attributing absurdity to our neighbour in certain untoward situations, but by and large homophony suffices. Yet there is nothing in either the neighbour's behaviour, or dispositions to behave, which obliges us to adopt this rule. There are no behavioural facts, so the argument goes, to prevent us adopting heterophonic translations whereby his 'is the same as' carries over into our 'belongs with' and his 'rabbit' into our 'undetached rabbit part', and so on across the language.

We can reconcile all this with our neighbour's verbal behaviour, by cunningly readjusting our translations of his various connecting predicates so as to compensate for the switch of ontology. In short, we can reproduce the inscrutability of reference at home.¹²

The final stage of the argument comes when we wonder about our own discourse and our own dispositions. The same points that were made concerning our interpretation of our neighbour's speech may be made mutatis mutandis when it comes to interpreting our own. As Quine says, "if there is really no fact of the matter, then the inscrutability of reference can be brought even closer to home than the neighbour's case; we can apply it to ourselves."¹³

The upshot is that there will be non-identical mappings onto itself of the infinite set of sentences of some one speaker's language which preserve invariant all his dispositions to assent to or dissent from sentences.¹⁴ As was the case in translating from a foreign language, those sentences onto which some one sentence is mapped will be different, though stimulus synonymous, because a single predicate will be mapped onto predicates that are (intuitively) not coextensive^{ve}. What is different, though, is that all of the predicates involved will be predicates of the same language. The theory of reference is at issue here, not just the theory of translation.

This development by Quine of his argument from translation challenges the very foundations of the realist position as I have explained it. If there is no determinate relation of reference between expressions of a language and parts of the world, then it might be doubted whether any of the four questions posed in the previous chapter can be answered. In consequence, one might be led to speculate on how much is left of realism when based on a weaker words/world relation such as might be distilled from Quine's notion of stimulus meaning. Such speculation, however, must surely signal a departure from the Fregean tradition. We can only make sense of saying what the objects of a theory are relative to some background language. But, if Quine is right, there are no behavioural facts which justify the preference of one assignment within a background language to others, and hence no fact of the matter as to which assignment is correct.

As Quine himself has observed, the example of rabbits and their parts and stages "is a contrived example and a perverse one, with which ... the practising linguist would have no patience."¹⁵ He goes on to present less bizarre cases that arise in practice. The clearest involves a claim about incompatible but equally acceptable translations of certain Japanese classifiers.¹⁶ Moving to the home language, others are said to arise when we use deferred ostension to establish some correspondence, "as in the case of a gasoline gauge."¹⁷ This suggests there might be an inscrutability in the choice between expressions and, say, their Gödel numbers as referents for quoted expressions. Finally, Quine has elsewhere expressed a guarded acceptance of Harman's example of the various referents of numerical expressions given by competing set theoretic reductions of number.¹⁸

Nevertheless, perversity is often the order of the day when it

comes to philosophical disputation. What is so important about the gavagai example and its development to include terms of our own language is that it opens up the possibility that every predicate that divides its reference does so inscrutably. The Japanese classifiers example involves inscrutability only in translation. As for the indecision about expressions and their Gödel numbers (also said to be a perverse example¹⁹), the ground is precisely the same as for inscrutability at home; we need to be able to readjust our translations of connecting predicates. One comes to realize that the vital step in each of these examples is the one summed up by Quine in this passage,

The whole notion of terms and their denotation is bound up with our own grammatical analysis of the sentences of our own language. It can be projected on the native language only as we settle what to count in the native language as analogues of our pronouns, identity, plurals, and related apparatus.²⁰

Quine finds this step plausible because, "of the broadly structural and contextual character of any considerations that could guide us to native translations of the English cluster of interrelated devices of individuation."²¹ Consequently, "there seem bound to be systematically very different choices, all of which do justice to all dispositions to verbal behaviour on the part of all concerned."²²

In section (iii) I shall argue that there are not the systematically different choices Quine thinks there seem bound to be. I shall concentrate on rather simple, but obviously central, cases of the gavagai kind. My aim is not to deny that there are no instances of inscrutability of extension for the predicates of a language, including predicates of our own language, English. It is rather to deny that there is the gross degree of inscrutability which Quine claims to have established, and which many philosophers take him to

have established, by reflecting on the relations between predicates and the grammatical machinery used in individuation. The conclusion I shall draw is that Quine has failed to show that there is not the determinate relation of reference assumed by the realist. There might be no fact of the matter as regards the relation for some particular term, but to establish this one needs to make an empirical investigation of the sort used by Quine in his discussion of the Japanese classifiers and the referents of numerical expressions.

In arguing thus I shall accept Quine's stricture that there is nothing more to guide us in translating or understanding a person's language than the facts about his behaviour and dispositions to behave. In Chapter 6, I shall argue that Quine overlooks, in his purview of translation, non-behavioural physical facts which might give us reason for preferring one translation to another. For the present, though, I shall confine my discussion to the behavioural. I shall also follow Quine in rejecting the myth of a museum. That is, I shall agree that for us to accept one translation or interpretation rather than another, there has to be a behavioural fact of the matter. The whole problem is to be precise about what these facts are.

The arguments I shall put forward derive support from some remarks made by Gareth Evans in a recent paper.²³ What Evans draws attention to is the fact that identity, and the whole notion of individuation, are in a strong sense secondary to predication from the point of view of a conceptual understanding of language. This suggests there is much more factual information than is considered by Quine. He is thus led to conclude that at least the predicates of our own language are not inscrutable — they do bear a determinate relation to the world. In the next section we shall see in what sense Evans thinks predication is a prior notion.

Section (ii): Evans on Identity and Predication

Evans's primary concern is with whether or not Quine has succeeded in showing, via the gavagai and rabbit examples, that there is "indeterminacy in the theory of meaning." That is, he is concerned with Quine's claim to have shown, by developing his argument from radical translation, that in any language, every predicate that divides its extension does so inscrutably. To this end he concentrates on certain primitive languages which may be said to constitute fragments of English.

"It is a quarrelsome man," says Evans, "who would bicker with Quine over the indeterminacy of translation — the constraints upon that enterprise being so slight."²⁴ It is far from clear, however, that the constraints on translation are so slight. In passing I mentioned Quine's availing himself of Wilson's "principle of charity" in order to avoid attributing absurdity to another's behaviour. It amounts to assuming that the beliefs of others, including foreigners, are much the same as our own in many commonplace matters. But belief is perhaps too shallow a category on its own. If our aim is to understand the behaviour, both verbal and physical, of a member of some language community, then it seems that what is required is not just a theory of translation but a theory of interpretation. We need to interpret actions — gestures, behaviour when prompted, etc. — before we can translate sentences. Both Davidson and Lewis have suggested that interpretation in this way underlies translation.²⁵ Clearly enough, a theory of interpretation requires a theory of action. Now in order to explain why a person acted in the way he did, it is not sufficient to simply attribute certain beliefs to him, e.g., about the outcome of his action, we also need to identify the desires he

has — why he wanted that outcome in the first place. Thus Richard Grandy urges more than charity,

If a translation tells us that the other person's beliefs and desires are connected in a way that is too bizarre for us to make sense of, then the translation is useless for our purposes. So we have, as a pragmatic constraint on translation, the condition that the imputed pattern of relations among beliefs, desires and the world be as similar to our own as possible. This principle I shall call the principle of humanity.²⁶

Lewis goes further still and suggests an additional five principles.²⁷ I shall discuss the matter further in Chapter 6.

Returning to Evans's article, it is clear that when he talks of indeterminacy he means that which might result from the inscrutability of foreign terms. As indicated at the end of the last section, on this point I am one of those who would bicker with Quine. By drawing attention to certain features of the structure and function of our own language, Evans presents strong reasons for rejecting Quine's claim to have shown indeterminacy in the concept of reference. In the next section I shall build on Evans's insight and offer a general argument against inscrutability both at home and as it affects our translation of a foreign language of the sort Quine considers. Not only is reference a determinate relation at home, it also makes sense to see it as determinate for a foreign language.

To begin with, Evans points out that Quine's argument for inscrutability rests upon the belief that, "the sole reason a semanticist can have for treating an expression as a predicate with a particular divided reference is to account for that expression's interaction with the (putative) apparatus of individuation."²⁸ We encountered this reason most recently in the passage quoted towards

the end of the last section which, I said, constitutes the vital step in Quine's examples. It originally arose in the context of Quine's explanation of "analytical hypotheses". Evans maintains that this reason is mistaken on two counts. To begin with, the "empirical location of the scheme of predication" is really to be identified by quite different means to its interaction with the apparatus of individuation. Furthermore, this apparatus itself only succeeds in locating the scheme because it is fixed in turn by these other means. Let us look at the first count in detail.

In a way reminiscent of some of Dummett's remarks on truth,²⁹ Evans argues that before we talk about reference we first have to be clear about the point of construing an expression, 'G', as a predicate which divides its reference over, say, rabbits and not rabbit parts or the like. The point is,

To explain how the truth conditions of certain elementary, but compound, sentences into which it enters are determined by their parts ... To see the notion 'what G is a predicate of' in this way is to see it as constrained by a theory of sentence composition into which it fits and which alone gives it sense.³⁰

Only in this way, he says, can the semantic theorist do justice to Davidson's insight that a learnable language must have a finite theory of meaning.³¹ Given this view about the point of interpreting a predicate in a particular way, the apparatus of individuation is indeed secondary. As an extreme case — discussed further in the next paragraph — we can imagine it being absent from a set of elementary sentences of the form 'Fx', where 'F' is some predicate and 'x' is a free variable. It will also no longer seem surprising that the notion 'what G is a predicate of' should then be underdetermined by data exclusively derived from the assent conditions of laconic one-word

utterances, like 'Gavagai', and of sentences putatively involving interaction with the apparatus of individuation, like 'Ip gavagai bleg op gavagai', "for by concentrating upon such data we thereby disregard precisely those compound sentences which give the notion its point."³²

What are the compound sentences that give the notion its point? By way of example, Evans asks us to consider a simple language in which the apparatus of individuation — plurals, pronouns, numerals, the identity predicate, the definite article — plays no part.³³ This language contains expressions G_1, G_2, \dots, G_n which, as one-word occasion sentences, we can establish as stimulus synonymous with our 'A rabbit!' (and hence with our 'A rabbit part!' and 'A rabbit stage!'), 'A man!', and so on. It also contains expressions F_1, F_2, \dots, F_n which, when queried, are assented to when the environment manifests the presence of certain general features that do not require the presence of a specific kind of object. They are stimulus synonymous with our 'White!', 'Furry!', 'Warm!', 'Bloodstained!', and so on. These expressions stand alone as occasion sentences and also with a sentential negation operator. (Quine accepts that we can determinately translate the negation operator.³⁴) It is also observed that complex sentences are formed by combining one of the F terms with one of the G terms, although two G terms are never coupled together. Furthermore, the negation operator is also found to occur with the F terms to yield an internal negation, $(\text{not-}F G)$, which is syntactically and behaviourally distinguishable from the external negation $\text{Not-}(F G)$.

Evans then argues that the assent conditions for $(F G)$, e.g., for 'White rabbit', may well turn out to be such that the F feature has to be distributed "in a characteristic way in relation to the boundaries of a single object whose presence prompts assent to the

queried G terms."³⁵ Neither assent to both F and G nor an overlap between the features associated with F and G is sufficient for assent to (F G). Also, in those cases where overlap is sufficient, as in 'Bloodstained rabbit', the assent conditions of the internally negated sentence (not-F G), "again show a sensitivity to the boundaries of an object, for assent requires the absence of the associated feature from the entire exposed surface of that object."³⁶

How are we to give a systematic account of the truth conditions of these and similar compound sentences? According to Evans, we first have to attribute to the parts of those sentences properties consistent with their occurrences in all other contexts, and then characterize the construction of those sentences in such a way that the truth conditions can be deduced. In section II of his paper, Evans explains in some detail what an acceptable account would look like. As is to be expected, it entails that 'rabbit' is to be interpreted as a predicate which divides its extension over rabbits. In section III, he argues that none of Quine's alternative semantic proposals are acceptable. Let me briefly summarize Evans's own account.

In so far as we are interested in constructing a theory of meaning for the simple language envisaged, we are obliged to state, amongst other things, how 'White' occurs in 'not-White'; the two expressions are clearly not associated with independent conditions. Suppose that the behavioural evidence warrants this general principle for generating the semantical contribution of 'not' $\cap \phi$ from that of ϕ : an object satisfies 'not' $\cap \phi$ if and only if the object does not satisfy ϕ . (Evans notes that Tarski showed this account unifies both internal and external uses of 'not'.) Then it follows that contradiction will result if both predicates are applied to the same object.

So the distribution of whiteness throughout a rabbit-shaped area and not some other is relevant to the judgments 'White rabbit' and 'not-White rabbit', "precisely because either judgment affirmed upon an insufficiently extended survey is liable to be contradicted by the other judgment, warranted by the condition of, and made with respect to, the same rabbit."³⁷

Here we touch upon what Evans calls "the deep connection" between predication and identity, and thus upon the rest of the apparatus of individuation. At this point, too, Evans makes clear the second count on which he thinks Quine's "sole reason" is mistaken. A necessary condition for a predicate to be the identity predicate is that the way speakers use sentences containing it reveals a disposition to withhold contradictory predicates from the things identified. Now there might well be cases where, because of the structure of the sentences of a language, we are forced, in giving a theory of meaning for the language, to recognize predication without also identifying an identity predicate. But for any language at all, we could never recognize an identity sentence save by its inferential connections to such predicative sentences; in particular, its use in conformity with the principle of the nonidentity of discernibles. Of course then, Quine was right to find the identification of the identity predicate underdetermined by the data he considered, "for upon such a basis one could not show that an expression behaves in the required way in relation to contradictory predicates."³⁸ But this underdetermination is relatively unimportant once we appreciate that identity has to be tied to the rest of the language. As Evans concludes, "we may suppose that what objects a language distinguishes and talks about is a matter embedded much deeper than Quine's talk of jiggling with the translation of the

individuating apparatus would lead us to believe."³⁹

This is as far as I wish to take the discussion of Evans's article. The line of his central argument should now be clear. The point of construing a predicate of a language as having a particular extension is to explain how the truth conditions of whole sentences of the language come to be determined by the references of their parts. We need to be able to explain this in order to satisfy the requirement that a learnable language has a finite theory of meaning. How to construe a predicate is what we focus on when we look at the inter-relations between, say, the negation operator and certain categories of expressions in the language. We are then seeking a systematic account of how the truth conditions of those sentences depend on their structure. This account requires conformity with certain principles of identity, but these are only recognized from the way in which the predicates of the language are themselves used. So explaining the scheme of predication of a language by considering the available behavioural evidence is epistemologically prior to investigating how predicates interact with the individuating apparatus. Thus, we can explain the truth conditions of compound sentences of, say, the form ($\underline{F} \underline{G}$) and ($\text{not-}\underline{F} \underline{G}$), given our recognition of their sensitivity to the identity conditions of rabbits, by suggesting that the sentences involve predicates of rabbits.

How successful is Evans in replying to Quine's claims to have shown that there is indeterminacy in the theory of reference? By drawing attention to certain features of the structure of our language and the way it functions, Evans is able to point to a mass of detail which has to be accounted for when we contemplate constructing a theory of meaning for our language. The onus is surely on Quine to

explain how his alternative proposals account for it in an equally satisfactory way. True enough, we have not looked at Evans's arguments against the possibility of such a Quinean retort. There are two reasons for this. The first is that my main purpose in relating Evans's proposals is to be able to draw upon them in the next section when I present my own arguments against inscrutability. The second is that, by replying to Quine's alternative proposals myself in the next section, the need for explaining Evans's replies is obviated.

As was noted at the beginning of this section, Evans sees little point in replying to the argument for inscrutability as it arises in translation from a foreign language. Thus he confines his attention to fragments of English, whose sentences have structures and assent conditions familiar to us. The "simple language" we considered was only a small fragment. In attempting to display shortcomings in Quine's semantic proposals, the fragments are considerably larger.⁴⁰ Obviously, the larger the fragment the less the arguments can be expected to carry over to the area of radical translation.

In the next section I shall be concentrating not on fragments of English, but on the domain of language Quine takes as the background for radical translation. I shall begin by making some further remarks about the structure of sentences containing the negation operator. The way in which the negation operator interacts with predicates will be contrasted with the way in which it interacts with singular terms. I shall then argue that the structural differences revealed in this way, together with other criteria if necessary, suggest a general way in which a linguist might identify a category of terms in a foreign language as corresponding to the category of singular terms in the home language, and thus refute an argument of

Quine's against the possibility of such an identification. Having identified this category, the linguist will then be able to recognize the foreign quantifier construction, which in turn will enable him to recognize logical similarities between the two languages. Given what Evans has said about the relations between identity and predication, together with one or two further remarks, it will be argued that no reason so far given by Quine shows that the linguist will not be in a position to uniquely translate the predicate which he at first naively takes to be that of identity. As we have seen both in this section and the last, once this is fixed so are the translations of the predicates.

It must be emphasised that I am not going to give some kind of generalization of Evans's argument. Evans places no weight on the reasons which a person constructing a theory of meaning might have for identifying certain expressions as singular terms. As I have remarked though, my arguments will have just as much bearing on the case where a linguist is trying to construct a translation manual for an alien language. Thus it should not be surprising that I will have to take into account evidence relating to other categories of expressions. Moreover, I will seek to show that, given the close connection between singular terms and quantifiers, it is to be expected that the category of expression which is of particular significance is that of singular terms.

Section (iii): Catching gavagai

Quine has contrasted how we can translate foreign predicates and quantifiers with how we translate the truth functions.⁴¹ The latter is facilitated by the fact that we can state "substantial

behavioural conditions" for interpreting foreign operators like negation and alternation.⁴² When it comes to quantified sentences, however, matters stand differently, for they "depend for their truth on the objects ... of which the component terms are true; and what those objects are is not uniquely determined by stimulus meaning. Indeed ... like plural endings and identity [they] are part of our own special apparatus of objective reference."⁴³ Moreover, we will first need to decide which foreign terms correspond to our predicates and which to our singular terms. (Of course, if there are no such classes of terms in the foreign language then there is no correspondence, but then nor will there be inscrutability.) According to Quine, though, even the distinction between predicates and singular terms is independent of stimulus meaning.⁴⁴ I wish to challenge this last claim.

In pursuing an inductive definition of the class of foreign sentences, a linguist will have recognized certain large classes of expressions which might be termed 'grammatical categories'. Will he have any reason to correlate one of these categories with our English category of singular terms? The problem he faces, says Quine, is similar to that involved in deciding on the extension of an expression like 'gavagai'. Thus, the singular term 'Bernard J. Ortcutt' "differs none in stimulus meaning from a general term true of each of the good dean's temporal segments, and none from a general term true of each of his spatial parts."⁴⁵ The only way to decide is by appeal to the apparatus of individuation. But is this in fact the case? What Quine is saying is that for a unique translation to be made the linguist has to accept one or more analytical hypotheses, for no appeal to the behavioural facts is in itself sufficient. Let us consider more closely, then, what some of these facts are.

To begin with, the linguist will be able to glean quite a lot of information by reflecting on the different ways in which singular terms and predicates interact with the negation operator. The difference is encapsulated in Aristotle's dictum that a quality has a contrary but a substance does not. This may be interpreted in the following way. To say that a quality has a contrary is to say that, for any predicate 'F', there is another predicate 'not-F' which is true of just those objects of which the original predicate is false, and false of just those objects of which the original predicate is true. To say that a substance does not have a contrary is to say that we cannot in general assume that, for any given object a, there is some object b, distinct from a, of which just those predicates which are false of a are true of b, and just those which are false of b are true of a.

The fact that objects lack contraries is of particular importance to the linguist. It is revealed in natural language by the refusal of a speaker to recognize the use of expressions of the form 'not-a'. For a speaker who uses 'Bernard J. Ortcutt' as a name, and hence as a singular term, the expression 'not-Bernard J. Ortcutt' is non-sensical. Consequently, the repeated use by an inquisitive linguist of 'not-Bernard J. Ortcutt' in attempting to produce a statement is always likely to result not so much in refusal to assent or dissent on the part of a native speaker, but in his expressing bewilderment. It is another matter when negation of a predicate is the putative issue. If the linguist thinks he can identify a foreign predicate 'F₁' and a foreign name 'a₁', then his repeated prompting of the native speaker with ('not' \cap F₁ a₁) should sometimes elicit either assent or dissent. Notice too that in talking of the expression of

bewilderment — an amazed look, a boggle — and the refusal to assent or dissent — a shrug, a gesture — we are dealing with features that are behaviourally detectable. Also, factual surprise or mere indifference in response to a well-formed sentence will not always occur; syntactic bewilderment in response to a deviant sentence always will.

The above remarks suggest the following guideline which might be used by a behaviourally conscious linguist in translating a foreign language:

- (A) For any (putative) singular term 'a' and predicate 'F' of a foreign language, if a user of that language either assents to or dissents from the sentence 'Fa', then he will usually express bewilderment when confronted by sequences of phonemes containing the expression 'not' \cap a, and will (thus) invariably refuse to assent to or dissent from them, though he will invariably either assent to or dissent from the sentence ('not' \cap F a).

Given his recognition that there are foreign grammatical categories, I submit that, on the basis of a sufficiently wide survey of the behavioural evidence, the linguist will be able to use (A) alone in order to settle on the foreign category most similar to our category of singular terms. It certainly decides the issue for the 'Bernard J. Ortcutt' example, and since this is Quine's only direct argument against the possibility of deciding on the class of singular terms, I conclude that he has failed to prove that the distinction between predicates and singular terms is independent of stimulus meaning.

We should further note that (A) is not the only guideline a linguist can use in attempting to detect singular terms. Leaving aside the foreigner's reaction to certain sequences of phonemes

containing the negation operator, suppose we look at those cases where he refuses to either assent to or dissent from an utterance which the linguist has good reason to believe is a member of the class of foreign sentences. It has been claimed by Strawson,⁴⁶ and accepted by Quine,⁴⁷ that this reflects a truth-value gap, i.e., a sentence which is neither true nor false. This might not always be the case. It might be, e.g., that assent or dissent would amount to a flagrant disregard for some social convention on the part of a speaker. Let us suppose, though, that such difficulties could be resolved by sufficiently varying the circumstances of the questioning.

Strawson goes on to point out that, "whether the sentence is true or false depends on the success or failure of the general term; but the failure of the singular term appears to deprive the general term of the chance of either success or failure."⁴⁸ This would seem to suggest, and Strawson certainly writes as though it suggests, that truth-value gaps will be seen to depend systematically on members of one grammatical category. Replying to Strawson, Quine claims that, "there will of course be the problem of deciding which word of the truth-valueless sentence to blame the truth-value gap on, and there will be other technical problems."⁴⁹ Unfortunately Quine does not elaborate on what these "technical problems" might be. The problem of "which word" seems to have more substance to it. Truth-value gaps arise in our own language for a variety of reasons other than failure of reference on the part of a singular term — category mistakes, presuppositions, ambiguity, vagueness, and the like. Can these be overcome by casting the net in search of evidence sufficiently wide? Surely we need to emphasise the empirical spirit here. And what is more, the linguist should be able, with the help of (A), to check any

hypothesis he may be led to form.

Another guideline is suggested by the evidence that can be deduced from reflecting on how singular terms interact with the category of predicates. In our own language, we never find a person assenting to, say, ' \underline{F}_1 Ortcutt' and ' \underline{F}_2 Ortcutt' without being prepared to assent to the compound ' $\underline{F}_1 \underline{F}_2$ Ortcutt'. Could the linguist use this as a further aid in translation? It would be a bizarre language indeed where a speaker assented to some object having property \underline{F}_1 and to the same object having property \underline{F}_2 , whilst refusing to assent to that object having both \underline{F}_1 and \underline{F}_2 ! As for the problem of identifying the foreign category of predicates, guideline (A) will be of help here. Notice also that such an aid casts further doubt on Quine's original argument against identifying foreign singular terms. If 'Ortcutt' were treated as a predicate true of several things, then the aid could be preserved upon one assumption only: that whatever is true of one Ortcutt is true of all. That is, if the principle is to hold, then the linguist can only regard what are putatively singular terms as predicates on the assumption that the several things of which the term is true are indiscernible by the predicates with which it interacts. This seems implausible.

As a final guideline to the establishment of some foreign category as the category of singular terms, let us once more look at how the negation operator functions in elementary sentences. Suppose that a speaker of the foreign language our linguist is trying to translate assents to a sentence the linguist takes to be of the form ' $\text{Not}' \cap (\underline{F}a)$, where ' \underline{a} ' is a name. He can check whether this is the form by seeing if the speaker also assents to ' $\text{'not}' \cap \underline{F} \underline{a}$ '. For if a person is disposed to assent to its not being the case that \underline{a} is \underline{F} ,

it would seem that he should be equally disposed to assent to its being the case that a is not-F (and vice-versa). Or at any rate, if someone maintains that a person might not be equally disposed, it is surely incumbent upon them to explain how the difference could be manifested via a speaker's dispositions to assent and dissent. The same points may be made mutatis mutandis for the case where the speaker dissents from 'Not' \cap (Fa). This suggests the following guideline for translation:

- (B) For any (putative) singular term 'a' and predicate 'F' of a foreign language, a speaker's dispositions to assent to or dissent from 'Not' \cap (Fa) are the same as his dispositions to assent to or dissent from ('not' \cap F a).

This completes the first stage of my argument against the inscrutability of reference. I have put forward several guidelines for helping the linguist decide which foreign grammatical category is most similar to our own category of singular terms. Each of them sought to exploit certain facts about the structure of natural language. The first enabled us to rebut Quine's contrary argument and I suggested that it offered sufficient evidence in itself for a decision to be made. I then suggested three other guidelines, of varying strengths, to which the linguist might also appeal. The second of these provided a further argument against Quine's claim. At no point in the discussion did I assume that the linguist requires more than the behavioural evidence for putting these guidelines into practice. There was no need to, the behavioural evidence itself is overwhelming.

The second stage of the argument concerns how the linguist is to recognize the foreign quantifier construction. This is not the

same thing as translating sentences of the foreign language containing quantifiers. For that we have to know, as Quine says, what things count as objects in the foreign language. How then does the linguist recognize quantification?

One answer emerges when Quine reflects on substitutional quantification,

Behavioural conditions for interpreting a native construction as existential substitutional quantification, then, are readily formulated. We fix on parts of the construction as candidates for the roles of quantifier and variable; then a condition of their fitness is that the natives be disposed to dissent from a whole quantified sentence when and only when disposed to dissent from each of the sentences obtainable by dropping the quantifier and substituting for the variable. A second condition is that the natives be disposed to assent to one of the sentences obtainable by dropping the quantifier and substituting for the variable.⁵⁰

This might seem to be a rather haphazard procedure since "for any one choice of native locutions as candidates for the role of quantifier and variable, an infinite lot of quantified sentences and substitution instances would have to be tested."⁵¹ But Quine's faith in the empirical method is great, "empirical induction is all we have to go on, and all we would ask."⁵²

There are as many kinds of substitutional quantification for a language as there are admissible substitution classes. Since the linguist is now able to recognize the foreign category of singular terms, though, let us focus on it as the substitution class. As Christopher Hill has pointed out,⁵³ the experiments described by Quine will sometimes indicate that the foreigners are using objectual, or referential, quantifiers. For if there is at least one existentially quantified sentence which commands assent, while each of its

substitution instances tends to provoke dissent, the foreigners will be counted referentialists, believers in nameless objects. Hill goes on,

On the other hand, when informants are found assenting and dissenting as Quine describes, their quantifiers can be viewed either way. Certainly they could be substitutional quantifiers. It is possible, though, that the natives have an ontology which is bigger than their stock of names, but that no predicate of their theory is satisfied by just the nameless objects. Since both hypotheses accord equally well with patterns of assent and dissent, Quine says that they come to the same thing: in this situation, he is unable "to distinguish objectively between referential quantification and a substitutional counterfeit" (Ontological Rel., p.67).⁵⁴

More recently, Quine has expressed some doubt about the claimed inability to distinguish,⁵⁵ but the intricacies of the argument need not concern us here. What is important to us is Quine's acknowledgement that there is a point to the question of whether a given foreign expression counts as an existential quantifier. The linguist can now add to his previous knowledge of the foreign truth-functions. He will be able to map foreign sentences onto those classes of English sentences which have the same logical form up to the level where substitutional and referential quantification diverge. To use a phrase of Hill's,⁵⁶ the linguist will be in a position to recognize "logical similarities" between the two languages.

We now come to the crucial third stage of the argument, where we need to bear in mind what Evans calls "the deep connection" between identity and predication. The linguist would like to be able to decide whether a particular foreign two-place predicate, say 'bleg', can be translated by our identity sign. What Evans's remarks suggest is that there is at least one criterion for deciding. This criterion is that the way speakers of the language use sentences containing

'bleg' must reveal a disposition to withhold contradictory predicates from the things identified; they must use it in conformity with the principle of the nonidentity of discernibles. We may formulate this principle as:

$$(ND) \quad (Fx \ \& \ \sim Fy) \supset \sim (x = y)$$

It is easily proved that (ND) is inter-derivable in first-order logic with the principle of the indiscernibility of identicals:

$$(II) \quad (x = y \ \& \ Fx) \supset Fy$$

Now in order to generate all the valid schemata of the logic of identity, given first-order logic, the axiom-schema (II) needs to be supplemented by the axiom-schema

$$(I) \quad x = x$$

This suggests a further criterion for deciding whether a foreign predicate corresponds to our identity predicate: speakers of the foreign language must show a disposition to assent to statements of self-identity.

To check whether the foreign two-place predicate 'bleg' is in fact the same as our identity predicate, the linguist can obtain what would in that case be instances of the axioms by using what he has established to be foreign singular terms. When a foreigner is found assenting to a sentence containing 'bleg' and two singular terms 'a₁' and 'a₂', and to a sentence containing 'a₁' (or 'a₂') in one or more of its argument places, then he must be disposed to assent to a third sentence of exactly the same form as the second but which has 'a₂' (or 'a₁') substituted one or more times for 'a₁' (or 'a₂'). Furthermore, for any foreign singular term 'a', a speaker must be disposed to assent to the sentence formed by substituting 'a' in both argument places of a sentence containing 'bleg'. It would seem, then,

that even within the Quinean strictures on translation, a linguist will be able to determine whether a foreign people share our concept of identity.

We now come to the fourth and final stage of the argument, where the linguist attempts to resolve the translation of 'gavagai'. In section (i) we saw that when the linguist uses a foreign expression like 'bleg' in an attempt to decide whether this gavagai is the same as that one, it seems that he has no way of telling, without recourse to voluntarily added maxims, if he is thereby asking about wholes, parts or stages. But having established (a) what to count as foreign singular terms, (b) logical similarities between the two languages, and (c) criteria for when to translate a foreign predicate as an identity predicate, it now seems that the problem will yield to research. Thus, suppose he is worried about whether he has asked if this undetached rabbit part is identical with that undetached rabbit part, rather than whether this rabbit is identical with that one. Then he will proceed as follows. To begin with he obtains a variegated animal, e.g., one with a white leg and the rest brown, or one to which he adds some dye on an appropriate part. Let us suppose he is lucky and finds a rabbit with a white leg. In the usual way he establishes which foreign word is stimulus synonymous with our 'white'; imagine the word is 'fisp'. On the basis of his knowledge of the logical similarities, the linguist then produces the foreign sentence most similar to

$$(1) (\exists x)(\exists y)(\text{gavagai}(x) \& \text{gavagai}(y) \& \text{fisp}(x) \& \sim \text{fisp}(y) \& \text{bleg}(x,y)).$$

Now (1) has the same logical form as both of the following:

$$(2) (\exists x)(\exists y)(\text{rabbit}(x) \& \text{rabbit}(y) \& \text{white}(x) \& \sim \text{white}(y) \& x = y)$$

$$(3) (\exists x)(\exists y)(\text{undetached rabbit part}(x) \& \text{undetached rabbit part}(y) \& \text{white}(x) \& \sim \text{white}(y) \& (x) \text{ belongs with } (y)).$$

The important difference, however, lies in the assent conditions under stimulation in the presence of the variegated rabbit. If the foreigner does assent to (1) then, given the truth of (ND), (2) would clearly be an incorrect translation. (3) is the obvious choice. In such a case, it follows not only that 'bleg' is to be translated as '... belongs with ____', but also that 'gavagai' is a predicate true of undetached parts of rabbits. If, on the other hand, the foreigner dissents from (1) then, of course, (3) is ruled out as the correct translation. 'Rabbit' might be the correct translation of 'gavagai', but so might 'rabbit stage'; we shall resolve this quandary shortly.

One possible reply to this line of argument would be that 'fisp' is not a simple predicate like 'white', but is a complex predicate meaning something like 'white in some place'. When it appears twice in one sentence, the place in question is assumed to have changed. The correct translation of (1), it might then be urged, would be

$$(2') (\exists x)(\exists y)(\text{rabbit}(x) \ \& \ \text{rabbit}(y) \ \& \ \text{white in some place}(x) \ \& \ \sim \text{white in another place}(y) \ \& \ x = y).$$

But even such a perverse hypothesis as this can be tested. If 'fisp' is really stimulus synonymous with 'white in some place', then for anything to be fisp there has to be some place in which it is white. That is, 'fisp' has to be understood as a two-place predicate. So whenever a foreigner assents to its being the case that there is some one thing that is fisp, he will also assent to a double instantiation of the form

$$(4) (\exists x)(\exists y)(\text{fisp}(x,y)).$$

Having resolved the problem about undetached rabbit parts, let us now see how the linguist copes with rabbit stages. The suggestion that the foreign people are more partial to stages of rabbits than to

persisting rabbits implies that no statement asserting the identity of gavagais over time will be true. In order to exploit this fact, our linguist will need to have some way of fixing the times of certain events — certain "ocular irradiation patterns" as Quine calls them — in a way acceptable to the foreigner. He should encounter little difficulty in this. If the foreigner is indeed partial to stages and talks about them in conformity with a concept of identity, he must be acutely aware of the passage of time. Perhaps the Quinean device of "properly timed blindfoldings" will allow him to settle on some temporal predications for sightings of gavagais. Maybe he will need to avail himself of other events, like a flash of lightning, or the setting of the sun, which he can use to establish a translation of 'before'. At any rate, there would appear to be numerous alternatives. Let us suppose that he decides the translation of our complex predicate 'seen before the lightning flash' is 'ky lan worrat'. Blending this with 'gavagai', 'bleg', existential quantifier constructions and a few other logical terms, the linguist produces a sentence which looks like

$$(5) (\exists x)(\exists y)(\text{gavagai}(x) \ \& \ \text{gavagai}(y) \ \& \ \text{ky lan worrat}(x) \ \& \ \sim \text{ky lan worrat}(y) \ \& \ \text{bleg}(x,y)).$$

Now (5) has the same logical form as both of the following:

- $$(6) (\exists x)(\exists y)(\text{rabbit}(x) \ \& \ \text{rabbit}(y) \ \& \ \text{seen before the lightning flash}(x) \ \& \ \sim \text{seen before the lightning flash}(y) \ \& \ x=y)$$
- $$(7) (\exists x)(\exists y)(\text{rabbit stage}(x) \ \& \ \text{rabbit stage}(y) \ \& \ \text{seen before the lightning flash}(x) \ \& \ \sim \text{seen before the lightning flash}(y) \ \& \ (x) \text{ is a stage of the same animal as } (y)).$$

But the difference between (6) and (7) is revealed in the same way as the difference between (2) and (3) — by looking at the assent conditions. In this case, though, the foreigner is not stimulated by

the presence of a variegated rabbit, but by the appearance of any one rabbit at different times. If he assents to (5) then, given the truth of (ND), (7) must be the correct translation. Consequently, the linguist would know to translate 'bleg' as '... is a stage of the same animal as ___', and 'gavagai' as 'rabbit stage'. Dissent from (5) on the part of the foreigner rules out (7) as a translation.

Over the last few pages I have been considering how to distinguish a foreign predicate true of rabbits from two perverse alternatives — one true of undetached rabbit parts and one true of temporal stages of rabbits — and how to distinguish the perverse alternatives from each other. I have shown how a cunning linguist can accomplish this by examining the ways in which the foreign predicate interacts with the foreign identity predicate. Where the foreigners are partial to parts and not to wholes or stages, they will assent to a statement, such as (1), to the effect that different parts belong to one and the same object; where they are not partial in this way they would take the statement as not being about parts at all, or as being about different objects. Where they are partial to stages and not to wholes or parts, the foreigners will invariably assent to a statement, such as (5), according to which the object identified changes with time; otherwise they would sometimes see the same object over time.

The problem does not quite end here. Throughout this chapter I have simplified the discussion by considering as possible translations of 'gavagai' only those expressions which divide their reference. When presenting the argument for radical translation in Word and Object, however, Quine also suggests as two other possible translations 'rabbithood' — the universal term — and 'the rabbit fusion' — the fusion, in Goodman's sense, of all rabbits.⁵⁷ The occasion sentence

'Gavagai' would then be stimulus synonymous with 'Rabbithood is manifested here' or 'The rabbit fusion is manifested here'. What is different about these alternative translations is that they are singular terms and so do not divide their reference at all. Evans considers these possibilities in the article I discussed; we must now do the same.

The obvious reply for us to make is that the guidelines presented at the beginning of this section for distinguishing the category of foreign singular terms will allow the linguist to settle the matter. Thus, there is no contrary of 'rabbithood' or 'the rabbit fusion'. If additional, independent evidence were required, the linguist could obtain it in the following way. Where there is indecision about 'rabbit' and 'rabbithood', he waits until the object which prompts 'Gavagai' moves. A predicate is found which applies to its first position but not to its second, and an appropriate variant of (1) is uttered to the foreigner. Assent indicates 'rabbithood', since any manifestation of it is a manifestation of the same thing, the universal 'rabbithood'. A similar procedure distinguishes 'rabbit' from 'the rabbit fusion'. Different rabbits have different predicates true of them, but a predicate true of "that single though discontinuous portion of the spatiotemporal world that consists of rabbits"⁵⁸ will not, indeed cannot, vary from rabbit to rabbit. Lastly, if the linguist is not sure whether 'rabbithood' or 'the rabbit fusion' is intended, he can make use of the fact that a detached part of a rabbit is a part of the discontinuous portion of the spatiotemporal world that consists of rabbits. Such a part is thus an instantiation of the rabbit fusion but obviously not of rabbithood.

In section (i) I reported an argument of Quine's which purported to show that, because of the inscrutability of reference, there could

be not only logically incompatible translation manuals from a foreign language to ours, but also logically incompatible assignments of references or extensions to many expressions of our own language. The leading ideas behind Quine's argument are that the theory of reference for a language — what the extensions of the predicates are and what the singular terms denote — can only be settled by examining the individuating apparatus of that language, and that this is underdetermined by all the behavioural evidence that could be collected because of the evidence's "broadly structural and contextual character." In section (ii) I closely examined what Evans thought was wrong with the first idea. Predication, not individuation, is the key notion intimately bound with reference. Individuation is a secondary notion and can only be explained by considering how predicates function in language. Evans's own conclusion is that there are not going to be the logically incompatible assignments of extensions to the predicates of our own language that Quine thinks there will be.

In this section I have taken the key points Evans made about identity and predication and shown that inscrutability does not arise in translation from a foreign language in the way predicted by Quine. Clearly, if there is not the predicted inscrutability here then the argument by analogy for inscrutability at home collapses. Throughout the discussion we have not exceeded the bounds Quine himself imposes on translation. On the basis of the predicted behavioural evidence we were able to refute his claims about the indeterminacy of translating from one grammatical category to another, and even within the grammatical categories of singular term and predicate we were able to resolve indecisions about how to translate the expression 'gavagai'.

All of this said, I do not pretend to have shown that all terms,

even within our own language, have a 'scrutable' reference. Thus, I have said nothing that casts doubt on cases of deferred ostension or on the example of the Japanese classifiers. I have confined my attention to attacking one argument which Quine hopes will establish inscrutability on a grand scale. It might turn out that an ingenious reply will save the purported example. The assumptions I have made about translation in this section, many of which are taken over from Quine to be used ad hominem, might be questioned, although they seem perfectly acceptable when construed as statements about how we interpret our own language. In the end, then, the charge stands that Quine's argument from radical translation pays far too little attention to the fine-grained structure of language, to the ways in which singular terms and predicates interact with truth-functions and quantifiers in order to give complex sentences whose truth-values depend on the references of their parts.

At the beginning of this chapter we noted that any statements about the extensions or referents of the expressions of a language are contained in a meta- or background language. Quine's argument purports to show that, for a very wide range of expressions, there will be logically incompatible assignments within the one metalanguage, and that there are no behavioural facts which would enable us to decide between them. I have argued that, on the contrary, there are behavioural facts which settle the matter for Quine's main example. So in this case there is a fact of the matter about which assignment is the correct one. My conclusion, therefore, is that Quine's argument fails to show that there is no determinate relation of reference between expressions of a language and parts of the world. In the next two chapters I shall explain what the significance of this relation is for the realist.

CHAPTER 4 THE FOUNDATIONS OF TRUTH

Section (i): Tarski's Theory as a Correspondence Theory

"Throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called classical conception of truth ('true — corresponding with reality')."

— A. Tarski, "The Concept of Truth in Formalized Languages"

At the very start of this thesis I distinguished two problems — the problem of reference, of to what extent scientific theories may be said to describe things that exist in the world, and the problem of predication, of whether scientific theories may be said to be true. The realist, I said, conceives of the theories of a mature science as giving approximately true accounts of how the world is. The instrumentalist, on the other hand, was said to commonly deny that this is so. For various reasons I have advocated accepting the realist's conception rather than the instrumentalist's.

When I first posed these problems, I drew attention to the fact that they were closely related to each other in the sense that one way of explaining what it is for any statement to be true is by showing how things in the world can be as it says they are. We might put this more strongly by saying that a statement is true precisely because the world is as the statement says it is. Here we have an expression of the correspondence theory of truth. Whether or not it is in fact a theory is a moot point; perhaps 'definition' would be better. My main concern in this chapter, though, will be with the term 'correspondence'. No capital will be made out of the choice to talk in the modern idiom of a theory.

The fundamental idea behind the correspondence theory of truth is that the property of being true is to be explained by a relation between a statement, or sentence of some interpreted language, and something else. This idea was explicitly recognized by Plato in his dialogue the Sophist.¹ It was later given its classical formulation by Aristotle in the Metaphysics: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true."² In various forms such a definition was discussed by Aquinas and Buridan in the later Middle Ages, and by Moore, Russell, Ramsey, and Wittgenstein in this century. My concern in this chapter will be with a particular variant of the theory due to Alfred Tarski.³

It is natural that a realist should accept some form of the correspondence theory of truth. Since he accepts that we are able to make true statements about things in the world, and that these things exist independently of both our sense experience and the theories we postulate to describe the world, what could be more natural than to say that the statements are true because of how these things are? The realist, moreover, does not see this ability to make true statements as something peculiar to us now; our forefathers were able to make true statements, true because they correctly described the world. This recalls the realist's conception of the growth of knowledge — our theories are simpler, have greater predictive success, and so on, than our ancestors', but in many cases they are theories of the same things.

Here the relativist disagrees. He does not deny that statements of a theory are true, and he might well not deny that their truth consists in saying how the world is. What he does deny, however, is

that there is a single external reality which is the measure of truth. We saw in Chapter 1 how Feyerabend talks about the ontologies of earlier theories being completely replaced by those of later ones, and of there being changes in world outlook. We also noted that Kuhn talks about the senses in which post-revolutionary scientists may be said to work in a different world from their predecessors. At the beginning of Chapter 2 it was pointed out that in order to say that two theories differ in point of world outlook, etc., it has to be assumed that they are comparable. This led us to interpret relativism as denying that we can establish that competing or successive theories are about the same things. It follows from this that there is no growth or convergence of knowledge in the realist's sense. It will in principle not be decidable if Dalton's "atoms" were Avogadro's "molecules"; when present-day physicists talk about electrons, there is no fact of the matter as to whether they are talking about what Bohr was referring to when he used the term 'electron'; Mendel's term 'bildungsfähig Element', which we translate as 'formative element', cannot be shown to have the same extension as our term 'gene'; and so on.

Some critics have gone further than this and have attempted to settle the realism/relativism debate on a priori grounds in the realist's favour. The relativist, they say, cannot even state his position without making use of concepts which are coherent only if realism is true. For the relativist, the truth-values of statements are relative to some sort of theoretical framework or to a set of core statements. The property of being true tends to be viewed as a three-place relation of a rather different sort to the two-place relation underlying the correspondence theory. But what about the truth-value of the statement that the truth-value of a statement is relative?

Is it also only relatively true, i.e., true relative to some further set of core statements? If this line of argument is followed through, it appears that the relativist is faced with an infinite regress, to be stopped only by tacitly accepting the realist's 'non-relative' notion of truth. I shall not, however, pursue this argument further.⁴ My whole approach in this thesis is not to attempt a refutation of relativism, but to respond to the challenge it poses to the realist to explain his view of the nature of scientific theories and the growth of scientific knowledge. The realist's view makes essential use of the notions of reference and truth, and one aim of this thesis is to firmly base these notions.

In saying that a statement is true because it correctly describes the world, the realist should not be taken as saying that truth must be explained in terms of a relation between a statement as a whole and some entity, perhaps a fact, or a state of affairs. This commonly results from interpreting the question 'What is it for a statement to be true?' as asking 'What makes a statement true?' Thus suppose we offer, as an answer to the former question, 'It is for the statement to correspond with the facts.' This immediately gives rise to two well-known objections. The first is that in order to understand the answer we must know what kind of correspondence is intended, and a concept such as that of a fact, or the facts, seems to stand in as much need of explanation as the concept of truth. Indeed, it appears that the only way in which we can explain what a fact is, is by saying that it is what a true statement states.

The second objection is that if anybody wants to assert that the relation between the statement and the fact does hold, he risks

sliding into a regress. For making the assertion would introduce a second-order correspondence. The first-order correspondence will either hold or not hold. Any question of its holding, if answered in the affirmative, amounts to the statement '"The correspondence holds" is true'. Now if truth is correspondence, this will introduce a second-order correspondence, and so on indefinitely. Dummett discusses this objection in his book on Frege.⁵ He argues that what it shows is not that truth is "absolutely indefinable", but that any legitimate definition has to meet a particular condition,

The condition is that the definition should yield the result that, e.g., to enquire whether the statement 'Frege died in 1925' is true is to enquire whether Frege died in 1925, and likewise for every other statement. That is, the infinite regress can be neutralized provided that the result of applying the definition of '...is true' to the specific instance '"Frege died in 1925" is true' is that this sentence is reduced to the sentence 'Frege died in 1925', and likewise for all other specific instances.⁶

Notice that truth is understood here as a predicate of sentences.

Dummett goes on to state the condition in its general form, "where A is any sentence, and S its canonical name, it should be possible to derive, from the definition of 'S is true', the equivalence 'S is true if and only if A'."⁷

Assuming that Dummett's argument is correct, our remarks over the last few pages suggest that, if any form of the correspondence theory of truth is to prove adequate, it must satisfy the following constraints:

- (i) it must explain the property of being true in terms of a relation between a statement and something else,
- (ii) this relation is not one holding between the statement as a

whole and some other kind of entity such as a fact, and
(iii) the condition stated by Dummett.

Let us now see how Tarski's theory satisfies these three constraints.

I have already introduced, in Chapter 2, several of the ideas behind the theory. To begin with, we talked about true sentences. In Tarski's theory, 'true' is treated as a predicate of sentences. The theory answers the question 'What is truth?' just in so far as the notion of truth coincides with that of 'true sentence'. This suits us well, for in asking whether or not scientific theories are true, we are asking whether or not their component sentences — the basic laws, and so on — are true.

In Chapter 2 also it was emphasised that a sentence is true only as part of some particular language; the same goes for a sentence being false.⁸ This point has been made by numerous commentators since the Middle Ages, and Tarski makes it too. He actually goes on to prove that a sentence asserting that some sentence S is a true sentence of some classical language L cannot itself be a sentence of L, but must belong to a metalanguage in which the sentences of L are not used but can be mentioned or discussed. The Aristotelian criterion takes no account of this and so leads to contradiction.

It is important to notice the qualification 'classical' used to describe the languages Tarski is concerned with. Kripke has recently outlined a theory of truth for certain non-classical languages, in particular those with truth-value gaps, which contain their own truth predicates.⁹ In this thesis I shall confine myself to classical languages and the standard, Tarskian theory of truth.

In sketching Tarski's theory, I will focus my attention on an object language L and a metalanguage ML, which are both built on the

pattern of first-order quantification theory. L will be assumed to have a paraphrase in some natural language, and ML a paraphrase in the natural language English. As well as being a quantificational language, L contains a finite number of names (a_1, a_2, \dots) and one-place predicates (F_1, F_2, \dots). Its syntax may be defined by giving recursive definitions of 'singular term' and 'formula', on the basis of which the 'closed sentences' are singled out.¹⁰ Closed sentences of the object language are the things that are true or false.

The most important feature of Tarski's approach is his adequacy condition, Convention T. ML must have structural-descriptive names of, and translations for, all closed sentences of L . Convention T requires of a theory of truth that, for each closed sentence of L , it have as a logical consequence in ML an instance of the schema

(T) X is true if and only if p

where ' X ' is replaced by a structural-descriptive name of the closed sentence of L , and ' p ' is the translation into ML of this same sentence. Clearly, Convention T is precisely the condition stated by Dummett as being necessary for any legitimate definition of truth. Thus, where L is a fragment of English, and the structural-descriptive name of the sentence 'Frege died in 1925' is taken to be formed by putting that very sentence in quotation marks, (T) yields as an instance

(T¹) 'Frege died in 1925' is true if and only if Frege died in 1925.

We shall see shortly how Tarski's theory meets his adequacy condition, but first a remark needs to be made that will be taken up later.

In (T¹) we have an instance of what has come to be known as the 'disquotation' principle: where L and ML are both fragments of the

same language, the name of the closed sentence of L is derived, as was said, by putting the sentence within quotation marks, and the translation of that sentence is taken to be the sentence itself, i.e., the translation is the name 'disquoted'. Some people have claimed, on the basis of this, that Tarski's theory is neutral with respect to epistemological questions. Putnam, for example, says, "On this view, 'true' is, amazingly, a philosophically neutral notion. 'True' is just a device for 'semantic ascent' — for 'raising' assertions from the 'object language' to the 'meta-language', and the device does not commit one epistemologically or metaphysically."¹¹ In the next section I shall contest this claim. Although instances of (T) may in this way appear neutral, this is quite different from saying that the theory is itself neutral. But first we must look further, at how the theory is established and at how, in general, instances of (T) are obtained.

The basis of Tarski's theory is the concept of satisfaction. The idea of the theory is to give a recursive definition of this concept, and then to connect it appropriately with the concept of truth. Satisfaction was conceived by Tarski as a two-place relation between an infinite numbered sequence of individuals and a sentence, open or closed, of the object language. I shall not give the details of the recursive definition here, but rather illustrate it for a part of L .

Suppose that, in L (again assumed to be a fragment of English), ' \underline{F}_1 ' is the predicate 'died in 1925'. Consider now the open sentence of L , ' $\underline{F}_1 \underline{x}$ '. This will be satisfied by a sequence \underline{s} if and only if the first member of \underline{s} died in 1925. The closed sentence ' $(\exists x) \underline{F}_1 \underline{x}$ ' will be satisfied by \underline{s} if and only if ' $\underline{F}_1 \underline{x}$ ' is satisfied by \underline{s} or by

some other sequence like \underline{s} except in having a different first term. Whether or not a particular sequence satisfies a sentence depends entirely on what individuals it assigns to the free variables of the sentence. So if the sentence has no free variables, i.e., if it is a closed sentence, then it must be satisfied by every sequence or by none. Thus, where it is assumed that Frege is one of the individuals in the domain of quantification of L , and that Frege died in 1925, ' $(\exists x)F_1x$ ' is satisfied by every sequence \underline{s} , because even if it is not the case that the first term of a given \underline{s} denotes an individual who died in 1925, there is a sequence exactly like \underline{s} except that its first term does denote Frege. Finally, it should be clear from these illustrations of the recursion that those closed sentences that are satisfied by all sequences are true, while those which are satisfied by none are false. This is how the sentential predicate 'is true' is defined — satisfaction by all sequences.

We now begin to see why Tarski's theory deserves to be called a correspondence theory of truth. What it is for a sentence of an interpreted language to be true is explained, non-trivially, in terms of a relation between the sentence and something else. This relation is captured in the concept of satisfaction. Moreover, the relation holds not between sentences and some other kind of entity, but between sequences of individuals and sentences of the object language. The first two constraints alluded to earlier are therefore met.

Before we turn to the third constraint, let us concentrate on the explanation of what is meant by saying of any predicate in the object language that it is satisfied by a given sequence of objects. It was said that ' F_1x ' is satisfied by a sequence \underline{s} if and only if the first member of \underline{s} died in 1925. Apparently this is only an

explanation in so far as the predicate ' \underline{F}_1 ' is itself understood. A similar point may be made with respect to the satisfaction condition for the closed sentence ' $(\exists x)\underline{F}_1x$ ', which requires that ' \underline{F}_1x ' be satisfied by \underline{s} or by some other sequence like \underline{s} except in having a different first term. To suppose that this condition explains the notion of existential quantification would therefore involve a vicious circle. Tarski, as Quine has noted,¹² saw the purpose of the condition the other way round; not as explaining existential quantification, but as contributing to a definition of satisfaction and so, derivatively, of truth.

In the case of the predicate, though, it appears that a natural way of attempting to avoid this restriction is by replacing the above explanation by, say, "' \underline{F}_1x ' is satisfied by a sequence \underline{s} if and only if ' \underline{F}_1 ' correctly applies to the first member of \underline{s} ". This, however, assumes that the metalanguage already contains the semantic expression 'correctly applies to'.

Now it was Tarski's intention to eliminate as many semantic concepts as possible from his investigations into the concept of truth. He says, "In this construction I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts."¹³ How he attempted to fulfil this intention is best seen for cases where names are involved. Suppose we wish to explain what is meant by saying that the closed sentence ' \underline{F}_1a_1 ' is true, where ' a_1 ' is the name 'Frege'. The question of satisfaction here turns on the question of what the name ' a_1 ' denotes. Consequently, a natural suggestion for explaining what it is for the sentence to be true would appear to be "' \underline{F}_1a_1 ' is true if and only if ' \underline{F}_1 ' correctly applies to the individual which ' a_1 ' denotes". This time, though,

the 'explanation' assumes not only that the metalanguage contains the semantic expression 'correctly applies to', but also the semantic expression 'denotes'.

Tarski attempted to provide a solution to these problems. It was to translate every name and predicate of the object language into English, and then utilize these clauses in the truth definition.

This strategy he states explicitly for 'denotes':

To say that the name x denotes a given object a is the same as to stipulate that the object a (or every sequence of which a is the corresponding term) satisfies a sentential function of a particular type. In colloquial language it would be a function which consists of three parts in the following order: a variable, the word 'is' and the given name x.¹⁴

Hence to say that 'Frege' denotes Frege is to stipulate that any sequence whose first member is Frege satisfies the open sentence (sentential function) 'x is Frege'. Given the way in which the satisfaction relation was defined, however, this strategy would appear to be circular, for what it means to say that the first member of a sequence satisfies the open sentence 'x is Frege' is that the first member is Frege. The assigning of objects to variables already presupposes the concept of denotation. An analogous remark holds for the suggestion that to say 'died in 1925' correctly applies to an individual a₁ is to stipulate that any sequence whose first member a₁ satisfies the open sentence 'x died in 1925'. For what it means to say that the first member of a sequence satisfies the open sentence 'x died in 1925' is that the first member died in 1925. A notion of correctly applying to is already presupposed. It seems as though in this case Tarski failed to see the purpose of assuming a prior familiarity with the notions of denoting and correctly applying to.

They too contribute to a definition of satisfaction and so, derivatively, of truth.

In the next section I shall consider an alternative way of dispensing with such an assumption, one which does not reintroduce the notion of satisfaction. It will be argued that only if the assumption can be dispensed with will it be correct to say, with Putnam, that 'true' is just a device for "semantic ascent". Meanwhile, let us leave Tarski's purported solution and accept that primitive notions of denoting and correctly applying to are presupposed in defining 'true in L'.

I said earlier that I would not present the whole theory for L and ML, but rather illustrate it for a part of some L. L was then assumed to be a fragment of English containing the sentence Dummett uses, 'Frege died in 1925'. We have seen how the theory satisfies Dummett's condition for this particular sentence. Obviously a full proof that the theory meets the condition for every closed sentence of L is out of the question without a full presentation of the theory. Nevertheless, as Tarski showed, when this is presented it meets his own adequacy condition which, as was noted, is equivalent to the one stated by Dummett. This is the third constraint we said had to be met by any form of the correspondence theory of truth which was to prove adequate. Furthermore, we have also seen how it meets the first two constraints. In so far, then, as these constraints are definitive of a correspondence theory of truth, Tarski's theory is such a theory.

From the realist's point of view, we may say that Tarski's theory is far more satisfactory than any theory which posits a relation between statements and facts could be. The question 'What is it for a statement, or sentence of an interpreted language, to be

true?' is answered in a way that makes essential reference to how the world is. This was seen most clearly when we looked at how the theory explains what it is for the statement 'Frege died in 1925' to be true. Any detour from explaining truth in this way, e.g., by first positing facts which correspond to a statement, must surely weaken the very notion of correspondence it is intended to support.

Section (ii): Truth and Reference

In this section I shall again consider that crucial part of Tarski's theory where a relation is established between names of the object language and individuals, and predicates of the object language and sets of individuals. What I want to do is to explore the connection between this and questions (2) and (3) of Chapter 2. These concern the epistemological and conceptual aspects of the theory of reference. But first I want to propose an answer to what I characterized as the most basic question of all, the question

- (4) What conditions have to be met in order for truth to be recursively defined?

In considering this question we need to remember that our attention is being confined to sentences of a classical first-order language, i.e., to a language suitable for the proof of all theorems of classical predicate logic, where the predicates range over properties of individuals and individual variables range over individuals. As we saw in the last section, Tarski showed how to define the concept of truth as it relates to the sentences of such a language. The question is, need the definition make use of any undefined relations?

The basic concept used by Tarski is that of satisfaction, which he defines recursively. The satisfaction relation carries the burden of reference, since only if there are individuals, and hence sequences of them, can we talk about the satisfaction of sentences by such sequences. There has to be a determinate relation of reference, then, in order for Tarski's definition to explain what the truth of a statement consists in. Reference in this sense underwrites truth, for the definition makes essential use of a relation between sentences of a language (words) and sequences of individuals (the world).¹⁵

Tarski thought that his notion of satisfaction could be used to define 'denotes' and 'correctly applies to' — the very notions that a theory of reference deals with. It seems, though, that doing so involves a vicious circle. But perhaps there is an alternative method of reduction, and if so this would cast doubt on the idea that a theory of truth presupposes a theory of reference. Suppose we say that names of the object language L primitively refer to their referents, and that predicates of L primitively refer to their extensions. Then it would seem that the relation of primitive reference could be defined by means of a list. Assume once more that L contains the one name 'Frege' and the one predicate 'died in 1925', then such a definition of primitive reference in L would be:

- (1) $(E)(x)(E \text{ primitively refers to } x \equiv E \text{ is 'Frege' and } x \text{ is Frege,}$
 $\text{or } E \text{ is 'died in 1925' and } x \text{ died in 1925}),$

where 'E' ranges over the names and predicates of L, and 'x' over the domain of L.

It should be noted that the relation of primitive reference is not introduced by Tarski. I am simply following Hartry Field in using a new term to focus on those semantical concepts of particular

interest to us.¹⁶ The question I now want to raise is whether this constitutes an adequate definition of primitive reference.

One person who has denied that it does is Field himself. He does concede that Tarski accomplished something of great philosophical importance in showing how to define truth, but he thinks that the definition only succeeds in explaining truth in terms of, or reducing truth to, a relation like that of primitive reference, and that a list takes us no further in understanding this basic notion. Here is the crucial paragraph from Field's paper,

Now, it would have been easy for a chemist, late in the last century, to have given a 'valence definition' of the following form:
 (3) $(\forall E)(\forall n)(E \text{ has valence } n \equiv E \text{ is potassium and } n \text{ is } +1, \text{ or } \dots$
 $E \text{ is sulphur and } n \text{ is } -2)$

where in the blanks go a list of similar clauses, one for each element. But, though this is an extensionally correct definition of valence, it would not have been an acceptable reduction; and had it turned out that nothing else was possible — had all efforts to explain valence in terms of the structural properties of atoms proved futile — scientists would have eventually had to decide either (a) to give up valence theory, or else (b) to replace the hypothesis of physioalism by another hypothesis (chemicalism?). It is part of scientific methodology to resist doing (b); and I also think it is part of scientific methodology to resist doing (a) as long as the notion of valence is serving the purposes for which it was designed (i.e., as long as it is proving useful in helping us characterize chemical compounds in terms of their valences). But the methodology is not to resist (a) and (b) by giving lists like (3); the methodology is to look for a real reduction. This is a methodology that has proved extremely fruitful in science, and I think we'd be crazy to give it up in linguistics. And I think we are giving up this fruitful methodology, unless we realize that we need to add theories of primitive reference to [the semantic definitions of the truth theory] if we are to establish the notion of truth as a physicalistically acceptable notion.¹⁷

Notice that (1), the definition of primitive reference for L by means of a list, is the linguistic counterpart of Field's "valence definition" (3). What Field is saying is that, since language is a natural phenomenon, we should apply, in our study of it, the same standards that we apply in other natural sciences; and this means we should not settle for (1). The relation of primitive reference is just as much part of the natural, physical order as the relation 'is chemically bonded to', and is to be studied in the same way. Both (1) and (3) are "extensionally correct" definitions — they correctly pair expressions with individuals, and elements with valences — but in order to be "physicalistically acceptable" they would have to give reductions of their definienda.

What sort of a definition does Field think is required before an acceptable reduction can be said to have been given? Evidently it has to be a causal-explanatory definition in which certain properties are explained in physical terms. In the case of valence, this would consist in relating the structural properties of atoms to the sorts of chemical combinations that the atoms, and their stable configurations (elements, radicals, etc.), enter into. This suggests that a 'recursive' definition of valence is in the offing, for "it is an important fact about valence that the valence of a configuration of elements is determined from the valences of the elements that make it up, and from the way they're put together."¹⁸

Field is less forthcoming about primitive reference. What would constitute a "real reduction" here? Perhaps an explanation in physical terms of why names and predicates have the references or extensions they do. This would provide an answer to my question (2) of Chapter 2. Field's only remarks on the matter relate to "primitive denotation".

He criticizes the "classical" view, attributed by him to Russell, that a name is "analytically linked" to a certain description, and conjectures that, when Tarski wrote, the only alternative seemed to be the trivial one encapsulated in the notion of a list. Then, in just half a paragraph, he outlines why he thinks a correct answer will have to take cognisance of the causal account of denotation. He does not think that Kripke or anyone else holds that purely causal accounts can be developed, even for proper names or natural kind predicates, but he does think that what they suggest is "a kind of factor involved in denotation that gives new hope to the idea of explaining the connection between language and the things it is about."¹⁹

Brief though Field's remarks are, I think he is right to draw attention to the causal account and to emphasise the need for a theory of primitive reference. Yet what he fails to do is to distinguish between the epistemological question of how we can discover what a name denotes or what the extension of a predicate is, and the conceptual question of what it is for a name to refer or for a predicate to have an extension. That is, he glosses over the distinction between questions (2) and (3) of Chapter 2. Field's analogy, and the implied requirement of a causal-explanatory notion of reference, obscure how an answer to (2) presupposes an answer to (3). Let me explain this further.

Field never says how an account of primitive reference is to be based on physical facts. What he does say is that "it seems likely that such things as psychological models of human beings and investigations of neurophysiology will be very relevant to discovering the mechanisms involved in reference."²⁰ But surely the mechanisms involved will relate to why a term has the reference or extension it

does, and not some other, rather than to what it is to have a reference or extension. It might well be conceded that psychology and neurophysiology will provide information about, e.g., the ways in which we group objects together using a single predicate, or how we go about individuating objects, but these are different matters.

It would seem that an explanation of what Field calls the concept of "primitive denotation" should explain such things as what conditions have to be satisfied by the name 'Frege' and the object Frege in order for our use of the name to denote the object. The fact that there are causal chains linking the object to uses of the name is of some importance. But how do we establish that there are these chains? And what is it that makes them causal? If we can trace them back then we might hope to discover which object is denoted. First of all, though, we need a criterion for uncovering and tracing them. At their base we imagine a more direct relation between the name, or particular events of its use, and the object. Information about the psychology and neurophysiology of the original users might tell us why the name denotes Frege and not something else. Yet it would seem to be an essential part of any explanation of how reference succeeds that the users intended to refer to Frege, and believed they were doing so. Causal relations are not basic; they too presuppose a theory of reference. These points will be discussed at length in the next chapter.

More recently, Putnam, on the basis of an idea of Leeds's, has argued that there are good reasons for not attempting to give a theory of primitive reference. He says, "we can give a 'transcendental argument' for Tarski's procedure by appealing to a purpose for having notions like truth and reference which is not at all parallel to the

purpose for which we have notions like valence." ²¹ Putnam earlier remarks that one purpose for having the notion of truth is so that we can state certain facts about deductive logic; another is said to be for expressing agreement when we do not know exactly what was said. Presumably he would want to say that one important purpose for having the notion of reference is so that we can define truth in terms of satisfaction. According to Putnam, it does not matter how these notions are defined, just so long as they do satisfy the purposes for which they are needed. If the list (1) can be shown to satisfy the purposes of having a definition of primitive reference, then it is acceptable. The crucial question now becomes, how can we be sure that it does satisfy those purposes?

When I sketched parts of Tarski's theory in section (i), I assumed, in order to facilitate the presentation, that the metalanguage ML was part of English. In contrast, the object language L was only assumed to be a part of some natural language. This is in accord with Tarski's procedure. Truth is defined for a given language which, in general, cannot be assumed to be that of the metalanguage in which the definition is stated. When it came to discussing examples, L was taken to be a fragment of English, although it was pointed out that this too was only done for simplification and that, in general, it has to be assumed, in deriving instances of Convention T, that there is some way of translating closed sentences of L into ML. Field expresses this point as follows,

. On Tarski's view we need to adequately translate the object language into the metalanguage in order to give an adequate theory of truth for the object language; this means that the notion of an adequate translation is employed in the methodology of giving truth theories, but it is not employed in the truth theories themselves. ²²

With this in mind, suppose that L is a quantificational fragment of German containing the name 'Deutschland', while ML is still part of English. Since the 'disquotation' principle will not apply here, we need to find some alternative translation on which to base our definition of primitive reference. Not just any translation will do. If, for example, 'Deutschland' were translated as 'Frege', a truth definition based on this would grossly misrepresent L, in which case the translation would not satisfy the purpose for which it was required.

On the basis of such a consideration, Field imposes, as an adequacy condition on the translation of names and predicates from object language to metalanguage, the condition of "coreferentiality": "two singular terms are coreferential if they denote the same thing; two predicate expressions are coreferential if they have the same extension, i.e., if they apply to the same things."²³ Putnam, however, sees the matter differently. He suggests that it is natural to say that the notion of reference only makes sense where the notion of translation makes sense, and hence that we should "think of reference as defined first for the home language à la Tarski, and then extended to other languages via translation."²⁴ But how can we be sure that disquotation, which is the way Putnam thinks Tarski defined reference for the home language, does satisfy the purpose for which it is required? Why is it that we are prepared to say that disquotation works for the home language? As far as I can see, Putnam has no answer to this question. If it is maintained that reference for the home language is defined using disquotation, i.e., that we do not need a theory of primitive reference, then there is nothing further to be said about how we know that the definition is adequate — either it is

or it is not, and that is the end of the matter. There are no epistemological questions that remain to be answered, and a theory of truth based on such a definition would be, in this sense, non-epistemic. All that there is to reference, on this view, is contained in the pairing of a term and an individual or a set of individuals, as in a list.

Of course, it still remains to be shown that disquotation is adequate as a means of establishing reference for the home language. In his paper, Field argues that it will not be when the language contains ambiguous names, counterfactuals, sentences in which 'denote' is applied to terms as they occur in more than one other language, or is such that it recognizes objects that cannot be named in the meta-language.²⁵ All of these features are present in English. I shall not pursue this objection here though.

If it is not possible to explain how we know that disquotation works for the home language, then ^{neither} ~~nor~~ is it possible to explain how reference works. As I have argued, we could not then even pose the question of how we can be sure that we do succeed in referring. This marks a serious defect with Putnam's proposal. But what if we were prepared to recognize the legitimacy of such a question? We should still have to explain the success of the disquotation principle, even though we did not accept it as a definition.

This can be done if we recognize the intuition underlying the causal theory of reference. For we can then hold that the principle works (to the extent it does) in virtue of a relation having been established between a word, or the event of its having been uttered in an appropriate circumstance, and an individual or a set of individuals. Our success in referring is grounded in the fact that

we stand in certain physical relations to what is being referred to. As was pointed out before, though, this still leaves unanswered the conceptual questions concerning the conditions which have to be satisfied by proper names and predicates in order for them to denote or correctly apply to objects.

There is also this point to be made against the disquotation principle as a definition. Tarski claimed that his definition of the concept of denotation using the satisfaction relation actually defined the meaning of the concept.²⁶ Now even if the disquotation principle were proved to be "extensionally adequate", in Field's sense, for even a part of the home language, this would still provide insufficient ground for claiming that the meaning of the concept of denotation was thereby defined. A list does not define the concept, or explain the meaning, of that of which it is a list. A list of valences does not explain what it is for an element to have a valence, and a list of individuals referred to does not explain what it is for a term to refer or to have an extension.

The main argument of this chapter has been that there has to be a primitive relation of reference between expressions of a language and parts of the world in order for us to explain, in the Tarskian manner, what the truth of a statement consists in. This suggests the following answer to question (4):

- (4A) There has to be a primitive relation of reference between
singular terms of a language and objects in the world, and
between predicates of the language and sets of objects in the
world.

Tarski's theory is a correspondence theory and is well suited to the realist. But it leaves unexplained what sort of a relation reference

is. How is it to be defined? What is needed is a theory of primitive reference, a theory of what it is for an individual to be the referent of a name, or to belong to the extension of a predicate, as those expressions are used within a community of speakers. This in turn should suggest how we can discover what the reference or extension of a particular term is.

CHAPTER 5 CLUSTER THEORIES OF REFERENCE

Section (i): Natural Kind Predicates and Proper Names

One central problem that I have frequently alluded to is that of explaining the growth of scientific knowledge. I have argued that the realist's account of this is intuitively plausible but requires a firm philosophical foundation; hence the four questions. According to the realist, many subsequent scientific theories give better accounts of just those things which earlier scientific theories described. I have mentioned several examples which might plausibly be interpreted in this way: Dalton's atomic theory was superseded by Avogadro's; the Bohr-Rutherford theory of the electron by present-day theories; Muller's theory of genes by the theories of modern molecular biologists. The realist does not claim that every change of theory is of this kind, for sometimes, as in the cases of 'phlogiston', 'luminiferous ether', and 'magnetic flux', we say that there is no one predicate now used which has the same extension as an earlier theoretical predicate. (One question we have yet to tackle is what makes these cases different.) All he is saying is that, by and large, scientists do, as a matter of fact, search for better theories about the kinds of things there are, and that science progresses because this methodology works.

The sorts of scientific predicates we are dealing with are often of a theoretical kind. As we saw in Chapter 1, there is no rigid distinction to be drawn between theoretical and observation terms. We may still use these expressions, however, for approximate description. Thus we might characterize 'atom', 'molecule', 'gene',

and 'electron' as theoretical. By contrast, some familiar predicates occurring in scientific theories might be characterized as observational, e.g., 'gold', 'water', 'quartz', and 'mammal'. These predicates can be correctly applied to objects of common experience. Correct application of theoretical predicates, on the other hand, depends on there being a certain amount of sophisticated apparatus. Water and gold are encountered in daily life in a sense in which genes and electrons are not. Also, correct application of predicates of an observational kind frequently requires little or no understanding of scientific theory.

There are other predicates, too, which are like observational scientific ones but which occur less frequently in scientific theories — predicates like 'lemon', 'tiger', 'oak tree', and so on. What these predicates have in common with predicates of the first two kinds, though, is that they are general names, characteristically associated with natural kinds. It is natural kind predicates that we are primarily concerned with.

There is no need for us to attempt to give a precise definition of what a natural kind is, but it is worth mentioning some features of them and of the predicates associated with them. To begin with, natural kind predicates do not usually admit of simple, precise definitions; as Putnam has said, they are "cluster terms".¹ 'Oak tree' might be defined as 'a member of the beech family with hard wood, jagged leaves, and which bears acorns', 'gold' as 'a yellow, non-rusting, malleable, ductile metal with atomic number 79'. These may be contrasted with 'bachelor', a general name not associated with a natural kind, which is commonly defined as 'a man who has never been married'. One interesting point which follows from this

difference is that whereas 'bachelor' cannot be correctly applied to an object that is either not a man or has been married, 'oak tree' and 'gold' can be applied to things that do not possess all the properties mentioned in their defining clusters; thus, many oak trees have leaves with smooth edges, and some gold is white. Putnam notes this point in saying that natural kind predicates are not synonymous with their definitions.² We might identify something as belonging to a kind even though it does not possess all the properties characteristically associated with members of the kind. This idea of natural kind predicates being cluster terms will be of great importance when we come to give a theory of reference for them.

A common feature of natural kinds is that the properties used to define them are primarily physical. Oak trees and gold are recognized by the physical properties they exhibit. Bachelors, on the other hand, are not to be distinguished from other men by physical properties but by a 'legal' property, viz., whether or not they have been married. Natural kinds are also commonly thought to be of explanatory importance and so become suitable subjects for scientific investigation. Consequently, many of the scientific predicates of interest to the realist, when he talks about scientific progress, are associated with natural kinds.

Having mentioned natural kinds, let me now dissociate myself from several of the things commonly said about them. In the first place, I do not wish to take the analogy between general names and proper names to the extent that a general name is held to refer to a kind of thing, i.e., to a supposed abstract entity, just as most proper names refer to particular things. There are electrons and sets of electrons, but not, in addition, a kind of thing 'electron'.

In the second place, I do not wish to commit myself to the modish view of Kripke and Putnam that natural kinds have essential properties, i.e., properties which nothing can lack and still be of the kind.³ More recent work has suggested that there are serious difficulties involved in trying to make sense of such a notion of essence.⁴ Towards the end of this chapter I shall suggest that we can make sense of a more restricted notion — that of "relative" or "epistemic" essentialism — but this, as Kripke emphasises, is not the sort of essentialism to which he is committed.

One last remark that needs to be made is that there is a set of terms whose members are frequently cited in discussions of the growth of science, but which are not associated with natural kinds — terms like 'temperature', 'mass', 'length', 'time', 'energy', and 'electric charge'. In scientific theories they typically express measurements, i.e., relations between objects, or events, and numbers, and so might be regarded as two-place predicates or relational terms. In the final chapter I shall consider how far the theory of reference developed in this chapter for natural kind predicates can be extended to include these terms.

Our primary concern in this chapter will be with answering the question:

(3) What conditions have to be satisfied by a natural kind

predicate 'Q' and an object a in order for 'Q', as it is used

within a linguistic community C, to be correctly applied to a?

An alternative way of putting this question would be, what is it for an object to belong to the extension of a natural kind predicate as that predicate is used within a linguistic community? One important point to notice about the question is that predicate extension is

relativized to use of the predicate within a linguistic community. It is clear that some restriction involving use has to be placed on an inquiry into what it is for a certain predicate to have an extension. Even though one linguistic community may use many of the predicates that another community uses, it does not follow that they are thereby talking about the same things. A mundane example arises when we reflect on the fact that whales were once thought to belong not to the natural kind associated with the term 'mammal' but to that associated with the term 'fish'. After reclassification, speakers still used the same terms, although their extensions had changed. Another example is suggested by the following passage, "The name 'electron' was introduced by G. Johnstone Stoney (1826-1911), in 1891, not of course as the name of the particles, but as the name of the fundamental unit of electricity, namely, the electric charge on a hydrogen ion in electrolysis."⁵

Given that the inquiry is to be made relative to use, the question is, whose use? The obvious choices are (a) use within a linguistic community, and (b) use by an individual speaker. I have chosen (a) because it is more germane to the enterprise of explaining how it is that subsequent theories can be about the same things. A scientific theory is something shared by a community, usually a community of scientists. Of course, this community is composed of individuals, many of whom know the theory. But often a speaker uses a scientific predicate, or even a relatively non-scientific natural kind term, without fully understanding it. This raises numerous problems for questions like (3) if use is taken as relative to an individual speaker; I will draw attention to some of them later. By choosing (a), I hope to avoid them.

In attempting to answer question (3) I shall make use of some of the ideas put forward in recent work on proper names. That this should be possible becomes apparent when it is recognized that there is a parallel question for proper names:

(3') What conditions have to be satisfied by a proper name ' α ' and an object a in order for ' α ', as it is used within a linguistic community C , to denote a ?

Or alternatively, what is it for an object to be the referent of a proper name as it is used within a linguistic community? I shall begin my investigation of (3) by considering how we might answer (3').

The theory of Frege's that I briefly outlined in Chapter 2 suggests one approach to answering (3'). It was pointed out that, in order to explain how certain identity statements could be informative, Frege was led to distinguish between the sense and the reference of a proper name. 'Sense' he took to be a cognitive notion, something which is known by users of the name and which can be passed on to others. In the case of proper names which occur in natural language, as opposed to proper names occurring in a perfect language, he thought that the sense may be different for different speakers,⁶ but for both these types of case the sense of a proper name would still be that by which the referent is determined. Dummett puts this point as follows,

What is important about Frege's theory is that a proper name, if it is to be considered as having a determinate sense, must have associated with it a specific criterion for recognizing a given object as the referent of the name; the referent of the name, if any, is whatever object satisfies that criterion.⁷

Before we can frame an answer to (3') on the basis of this theory, we need to decide what the sense of a proper name is as that

name is used within a linguistic community. Ought we to interpret it as the intersection of the senses associated with the name by the members of the community, as the union of those senses, or as something more complex? Of course, the answer to (3') would be straightforward if we had phrased it in terms of an individual speaker's use of a name: 'x' would denote a if and only if a were that object which uniquely satisfied the criterion for recognition associated with the sense attached to 'x' by the speaker.

One point that has to be taken into account here is captured by Putnam in his "principle of the division of linguistic labour", i.e., the principle that,

Every linguistic community ... possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets.⁸

As regards proper names, an example where Putnam's point is made clear is in our use of the name 'Einstein', where we intend to refer to the famous scientist who first formulated the Special Theory of Relativity, and so on. Many of those who use the name with this intention may not associate with it a criterion for recognizing the man; at best they only have a partial grasp of the sense. Nevertheless, there are dictionaries, encyclopedias, and more knowledgeable people who could provide the requisite information, and the realization that such sources may be appealed to grounds the use of the name within the community. Thus it would be a mistake to identify 'sense within a community' with 'intersection of senses for members of the community'.

Putnam's principle might be further extended if it is thought that the criteria associated with a name might not even be fully known

to any of the speakers. It might be that certain beliefs and items of knowledge regarding the bearer of a name are preserved in manuscripts and historical documents such as birth registers without anyone being aware of their existence, and it could be said that these were constitutive of the sense of the expression as it is used within a community whose members had access to the manuscripts and documents. In such cases we would have to say that 'sense within a community' was something more than 'union of senses for members of the community'. To allow for such a possibility, I shall henceforth understand 'the sense of a proper name as that name is used within a linguistic community C' as being given by a complex set of descriptions which may be arrived at on the basis of what is believed within C. 'What is believed within C' is, in turn, to be understood as including 'what is known within C'. Seen in this light, sense is a function of belief, and what is believed within a community may exceed what is believed by its members.

This stipulation means that some allowance will have to be made for inconsistencies within a community. If inconsistent descriptions are attributed to the bearer of a name, then clearly only one at most can be true of the bearer. Consequently, if the sense is to provide a criterion for determining the reference, only one of the descriptions can be reckoned part of the sense.

This theory which we have developed might seem somewhat removed from Frege's pristine view. Sense is to be thought of as a complex notion which cannot be assumed to be given by a simple definite description and might change over time. We have, however, preserved what was crucial to Frege's view — sense is a cognitive notion which, in the case of a proper name, provides a criterion for determining the referent.

We are now in a position to give a first answer to question (3'). It might be loosely termed 'Fregean':

(3'F) A proper name 'x', as it is used within a linguistic community C, denotes an object a if and only if a is that object which uniquely satisfies all of the descriptions which could be consistently attributed to the bearer of 'x' on the basis of what is believed within C.

Both the theory of reference encapsulated in (3'F) and Frege's original theory are examples of what have come to be known as descriptive theories of reference for proper names. They appeal only to beliefs and items of knowledge in order to determine what a name refers to. Other descriptive theories have been proposed by Wittgenstein,⁹ Searle,¹⁰ and Strawson.¹¹ Like (3'F) they take the sense of a proper name to be given by a cluster of descriptions; unlike (3'F) they do not maintain that the referent has to satisfy all of the cluster. Rather, the object named is that which satisfies a suitable number of them. A suitable number need not be a majority — allowance can be made for attaching more importance to some than to others. Consider, for example, our use of the name 'Archimedes'. We believe a number of things to be true of Archimedes — that he discovered the principle named after him, that he lived most of his life in Syracuse, that he once leapt from his bath and ran naked through the streets shouting 'Eureka!', that he invented a mechanical screw used for irrigation, and so on. Not all of these beliefs need be assumed to carry equal weight in determining who 'Archimedes' refers to. We might well consider satisfaction of, say, the first and second beliefs to be of primary importance.

There is good reason to revise (3'F) in order to account for

this view. If historical research should suggest, e.g., that one Anaximedes really invented what we call the 'Archimedean screw', then we would normally conclude that this is something Archimedes did not do. Sticking to the letter of (3'F), though, we would be forced to conclude, on the basis of this one piece of information, that there was no such person as Archimedes, and this seems most implausible. More generally, there seems to be no reason for holding that a person has to have done everything we attribute to them before they can be said to be referred to when we use their name. After all, we often concede that we have made a mistake in our description of a person, even when that description embodies the best knowledge available to us. This idea of a weighted cluster also marks a further analogy between proper names and natural kind predicates. Let us reformulate (3'F) so as to take account of it:

(3'C) A proper name ' α ', as it is used within a linguistic community C, denotes an object a if and only if a is that object which uniquely satisfies a suitable majority of the descriptions which could be consistently attributed to the bearer of ' α ' on the basis of what is believed within C.

One of the important features of descriptive theories is that the referent of a proper name is fixed by certain beliefs. In the case of 'Archimedes', the referent, for a linguistic community C, is that person (if any) who satisfies a suitable majority of those descriptions which could be attributed to him on the basis of what is believed within C. If it should turn out that one Anaximedes is so identified then 'Archimedes', as used within the community, denotes Anaximedes, despite the fact that there may have been some other person, contemporaneous with Anaximedes, who was named 'Archimedes'.

An illustration involving a case like this is afforded by the Gilbert and Sullivan operetta H.M.S. Pinafore. Towards the end, Buttercup, a nurse and child-minder, confesses that some years previously she mixed up two babies. One, Ralph, grew up to be an ordinary seaman, while the other, Corcoran, became captain of the Pinafore. According to a descriptive theory, it could be said after the confession that, in using the name 'Corcoran', people prior to the confession thereby succeeded in referring to Ralph.

Another important feature of descriptive theories, at least as I have presented them, is that, since the sense or meaning of a proper name is given by the same set of descriptions used to fix the referent, it would be meaningless to deny that the referent satisfied those descriptions. If the sense of the name 'Archimedes' is given by the four descriptive phrases mentioned above, then the statement 'Archimedes did not discover the principle named after him, nor live most of his life in Syracuse, nor run naked through the streets shouting 'Eureka!', nor invent a mechanical water-screw' would be contradictory. Conversely, we can say that the statement that he did do a suitable majority of these things is necessarily true.

These and other consequences of descriptive theories have been subject to stringent criticism by Kripke in his paper "Naming and Necessity".¹² According to Kripke, it is a mistake to think that such theories either fix the reference of a proper name or give its meaning. He argues that what one must do, in order to understand which object is referred to by a name, is to look at how that name came to be used in the linguistic community. In the next section we shall consider his criticisms in detail. My main argument will be that what he succeeds in drawing attention to is a feature of the use

of proper names which is relevant to the epistemological question 'How can we discover what the referent of a proper name is as that name is used within a linguistic community?', but not to the conceptual question of what reference consists in. That is, he draws attention to a feature relevant to answering the proper name version of question (2). Kripke has no theory of reference to offer, despite what others might claim. All he gives is an account of how we might decide, in certain cases where (3'C) proves inadequate, what is being referred to.

Section (ii): Kripke's Remarks on Naming

To be fair to Kripke, he does not explicitly mention the descriptive theory embodied in (3'C). His criticisms relate primarily to those theories in which the reference of a proper name is made relative to a speaker's use of the name. Against these he employs several examples, like the 'Einstein' one given above, which appear to be conclusive.¹³ Some of his criticisms, though, are more general, and it is clear that he thinks they are telling against any theory which implies that the referent of a proper name is whoever or whatever satisfies a certain set of descriptions.

One point which Kripke makes much of is that a proper name may be introduced via a definite description without its thereby coming to be held as synonymous with it.¹⁴ There is no reason why this point cannot be accommodated by one who holds a cluster version of the descriptive theory, such as (3'C). In fact, recalling Putnam's point about cluster concepts, he would seem to be committed to it. The name 'Mars' may have been introduced via the definite description 'the red star', but as we learn more and more about the planet so

the sense of the name changes.

Kripke then goes on to offer a general argument for why the sense of a proper name could never coincide with that of a definite description: names behave differently from definite descriptions in modal contexts. An extensive reply to this argument has been given by Dummett.¹⁵ Since proper names are being discussed in this chapter as something of a means to an end, it is not appropriate to involve ourselves in the details of what Dummett says. His main conclusion is, "For modal contexts in general, there is no relevant difference between proper names and definite descriptions."¹⁶ The one qualification he adds is that this is not so "when the name or the description is preceded by the verb 'to be' or 'to become'."¹⁷ But the reason for this "is not a general feature of the behaviour of proper names in modal contexts"; it rather has to do with the fact that a property like that of 'being Archimedes' is not a property that can be acquired. Thus, the person who did most of those things attributed to Archimedes did not become Archimedes when he did them; he had always been Archimedes, because he had always been the one who in fact was to do most of those things.¹⁸ One might question here the propriety of speaking of Archimedes as 'the one who was to do most of what is attributed to Archimedes'. Is it not possible that Archimedes might not have done any of those things attributed to him? This is in fact another argument which Kripke uses to cast doubt on descriptive theories; we shall examine it shortly.

What is more relevant from the point of view of the analogy we have drawn between proper names and natural kind predicates, and our attempt to throw some light on what it is for a natural kind predicate to have an extension, is what Kripke says about how the reference

of a proper name is fixed. Against the descriptive theorist's account of this, Kripke presents the following purported counter-example,

Imagine the following blatantly fictional situation ... Suppose that Gödel was not in fact the author of [the incompleteness] theorem. A man named 'Schmidt', whose body was found in Vienna under mysterious circumstances many years ago, actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the view in question, then, ... we, when we talk about 'Gödel', are in fact always referring to Schmidt.¹⁹

Kripke thinks that such a conclusion is quite wrong, and he is right to think so. But he is mistaken if he thinks that a descriptive theorist must be committed to it. What makes a counterfactual situation like the one given intelligible is that the sense of the name 'Gödel' is given by a cluster of descriptions which together provide a criterion for determining the referent. Schmidt would have had to have done more than to have first proved the incompleteness theorem in order for him to be the one to whom we refer when we use the name 'Gödel'. One might of course consider the case where the only thing that is believed about Gödel is that he was the first to prove the incompleteness theorem. Then a question would arise, given Kripke's fictional situation, as to whether users of the name had a false belief about Gödel or a true belief about Schmidt. Presumably Kripke would take the former to be the case.²⁰ The problem with this view, however, is that it renders unanswerable the question 'Who, then, is this Gödel about whom the users of the name had a false belief?'. For obviously any answer, even 'The man who was named 'Gödel'', would provide a criterion other than that associated with the original belief for recognizing the referent of the name 'Gödel'. In sum, Kripke's example is only telling against one who holds

strictly to a Fregean theory like (3'F); it has no force against one who holds a cluster of descriptions theory like (3'C).

Perhaps the most important example which Kripke gives concerns the name 'Aristotle'. It is similar to the Gödel one in that Kripke considers it to be telling against the descriptive theorist's account of how the referent of a proper name is fixed. But it goes beyond the Gödel one in attempting to cast doubt on what the descriptive theorist thinks the sense of a proper name is. Here is the key passage:

Not only is it true of the man Aristotle that he might not have gone into pedagogy; it is also true that we use the term 'Aristotle' in such a way that, in thinking of a counterfactual situation in which Aristotle didn't go into any of the fields and do any of the achievements we commonly attribute to him, still we would say that was a situation in which Aristotle did not do these things.²¹

Despite Kripke's claims to the contrary, it would be wrong to construe this argument as telling against the descriptive theorist's view that the referent of a proper name is fixed by certain beliefs. If we take the name 'Aristotle' and successively deny the truth of the usual descriptions given of him, there soon comes a point where the natural reaction is to say 'Who are you now talking about?'. Thus, if someone says 'Aristotle did not study with Plato, nor teach Alexander the Great, nor write the Metaphysics, nor ...', one soon wonders whom is being referred to. In reply to this, it might be said that what Kripke really meant to draw attention to, regardless of what he actually said, is the modal statement 'Aristotle might not have studied with Plato, nor taught Alexander the Great, nor written the Metaphysics, nor ...'. However, if this is understood

as 'Aristotle might not have studied with Plato, and Aristotle might not have taught Alexander the Great, and ...' then, although quite intelligible, it does not support the thesis that the referent of 'Aristotle' is not fixed by at least some set of descriptions. If, on the other hand, it is understood as 'It might have been the case that (Aristotle did not study with Plato, nor teach Alexander the Great, nor ...)', the embedded sentence is simply the one originally found to be puzzling.

Kripke's argument is more plausibly construed as an argument against the view that the meaning of a name is given by a certain set of descriptions. We noted in the last section that, given (3'C), it is necessarily true that the referent of a proper name satisfies a suitable majority of the descriptions associated with the name. From this it follows that there should come a point where a conjunctive assertion like 'Aristotle did not study with Plato, nor teach Alexander the Great, nor ...' not only invites a query but actually results in contradiction.

It is interesting to note here that Kripke himself provides an argument that could, in some instances, be used to bolster the descriptive theorist's case. For Kripke, as was noted in section (i), also wants to hold that objects have essential properties, where an essential property of an object is one which it could not have failed to have.²² The examples he gives of such properties involve their origin and substance.²³ He argues, for example, that it is necessarily true that a person has the parents they do. Hence, if it were known that, say, Aristotle's parents were X and Y, it would be contradictory to assert of Aristotle that his parents might not have been X and Y.

It does not follow from this of course that one who supports a descriptive theory need accept such properties as essential, or, indeed, that they need be committed to any form of Kripkean essentialism. What does seem to be inescapable though is that some more or less complex description is necessarily true of whoever or whatever is the bearer of a proper name. Or at any rate this will be so unless it is maintained that a cluster of descriptions determines the referent of a name but does not give its meaning. One who holds this last position would seem in turn to be committed to denying that names have meaning. For it is difficult to see that anything could be made of the view that names do have meaning, that what fixes the referent of a name is a criterion associated with what is believed about the bearer, but that the meaning of a name is not given by what is believed about the bearer.

I shall not discuss in detail that position which does not accord names meaning, for it marks a radical departure from the Fregean tradition. The meaning of a word, according to Frege, is just what a person knows when they know how to use the word, and so to deny that names have meaning seems to render unintelligible what a person's grasp of the use of a name consists in. If this criticism holds good, then one who wants to adopt a broadly Fregean view will have to hold that a cluster of descriptions not only fixes the referent of a name but also gives its meaning.

Returning to (3'C), we can say that one who wished to accept this account of how the sense or meaning of a proper name is to be specified would be committed to holding that it is necessarily true that the object which uniquely satisfies a suitable majority of the descriptions which could be consistently attributed to the bearer

of a name on the basis of what is believed within a linguistic community C, does satisfy some subset of those descriptions. Such a consequence of holding a descriptive theory might at first seem unpalatable. What needs to be borne in mind, though, is that the complex description which is said to be necessarily true of the bearer of the name is dependent on what is known and believed within the linguistic community in which the name is used. There is no imputation of necessary properties in the sense which Kripke calls metaphysical and which underlies his form of essentialism;²⁴ the kind of necessity is purely epistemological. The descriptive theorist is not in the position of holding that certain things will be true of the bearer of a name come what may, but only that they are true relative to what is known and believed within some linguistic community in which the name is used. In our own case, given those descriptions which we now believe to be true of Aristotle, it is not possible that a suitable majority of them should turn out not to be true of Aristotle. The suggestion that this might be so is tantamount to suggesting that the name 'Aristotle' is being analyzed as it is used within a different linguistic community from our own. The only sense which we can attach to the name, the only sense it has for us, depends on how we use it, and this is revealed by what we are prepared to believe about Aristotle. As we shall see in section (iii), this notion of 'epistemic essentialism' can be made out for natural kind predicates too.

So far in this section we have looked only at Kripke's negative remarks about the theory of naming and not at his positive ones. I have argued that even though his arguments may have some force against certain descriptive theories, they have very little against

the theory contained in (3'C). In particular, they certainly do not tell against (3'C) construed as a theory of how the referent of a proper name is to be determined. Kripke does have his own account of how the referent of a proper name is to be determined.

Furthermore, although he denies that names have meaning or sense of the cognitive kind that we have been talking about, he does make use of a function which he sees as performing a similar role in some contexts. What we now have to do is to see whether these views add anything to the theory of (3'C) about what it is for a proper name to refer.

As I have emphasised already, our main concern is with understanding what determines the referent of a name, for we eventually want to cast light on how we can decide what natural kind predicates have as their extensions. I shall therefore have little to say about Kripke's notion of meaning. The picture he presents in "Naming and Necessity" is one in which the meaning of a singular or general term is a function from possible worlds to objects or sets of objects. Thus, the meaning of the proper name 'Archimedes' is a partial function which assigns, to each possible world for which it is defined, the object which, in that possible world, is the referent of the name. The problem with this view is that although it might be technically useful in the semantics of modal logic, it in no way elucidates our understanding of what the meaning of a name is. Before we can check to see which object is assigned to 'Archimedes' in world w_1 , we will have to decide whether or not w_1 is a possible world relative to this one. How are we to decide this? Evidently it will depend on what counts as a necessary truth, i.e., on what is stipulated as true in all possible worlds. But what this means

is that we will first have to decide such things as whether or not a particular attribution to 'Archimedes' is possible, and the only way to do this is by reflecting on what meaning, in the cognitive sense of that term, the name 'Archimedes' has. In order to decide whether or not 'Archimedes' might have referred to some particular individual, we will need to reflect on what can truly be said of Archimedes, and this presupposes that we already have an account of the descriptive, cognitive kind of what the meaning of the name is.

Turning to what Kripke says about how the referent of a name is to be determined, it must be noted that Kripke himself denies that he is giving necessary and sufficient conditions for when a name can be said to refer to a particular object. His causal account, unlike the descriptive theories of the last section, offers no way of eliminating the notion of reference. To see why, we need to have an outline of the causal account before us. As Kripke observes, it is difficult to state precisely, but for the purposes of discussion I shall assume as standard some such account as the following: there is a name-giving ceremony in which the object is named by ostension, or the referent fixed by a description. Subsequent speakers then intend to use the name to refer to the same object. In this way a causal network develops, and in order to determine what a name refers to, as it is used in some community of speakers, one has to trace back through the network to the original ceremony.²⁵

The inclusion in this account of the condition that subsequent users of a name must intend to refer to the same object means that the notion of reference cannot be eliminated. The reason for including the condition is straightforward. To take an example of Kripke's, if I call my pet aardvark 'Napoleon', this may be causally

connected with the use of the name for the Emperor, but should lead to no identification of the Emperor with my aardvark.²⁶

Dummett has criticized Kripke's causal account on two grounds.²⁷ The first is that it fails to explain what the required intention is an intention to do: "Each speaker must intend to refer to the same object as the speaker from whom he heard the name: but what is it to refer to an object?"²⁸ Evidently an answer to this question is required before we can explain what it is that a speaker must intend to do in order that his use of a name count as another link in the causal network. Descriptive theories provide answers to this question and so it would seem that they are theories of reference in a sense in which Kripke's causal account is not.

The second objection Dummett raises is that it seems inevitable that there will be cases where the intention to preserve reference is fulfilled, but even so the referent is unwittingly transferred. In such cases it would not be possible, even in principle, to trace the causal network back in the way suggested by Kripke's account. An actual example where this appears to have happened is given by Evans in a recent paper.²⁹ He quotes from Isaac Taylor's book, Names and their History, "In the case of 'Madagascar' a hearsay report of Malay or Arab sailors misunderstood by Marco Polo ... has had the effect of transferring a corrupt form of the name of a portion of the African mainland to the great African Island." Such a situation is perfectly intelligible to us because we have a criterion, although imprecise, for 'the same geographical area', and we have a criterion for determining what geographical area a place name is being used as the name of, independently of the actual origin of the name. A cluster theory like (3'C) explains what the

latter criterion is. As Dummett remarks, though, Kripke's causal account "leaves no room for the occurrence of a misunderstanding: since to speak of a misunderstanding would presuppose that the name did in fact have a sense which could be misunderstood."³⁰ Consequently it is unable to make intelligible to us how what is now a name for an island was once a name for a part of the mainland. Moreover, it also fails as an account of what the referent of a name is because a causal network of the kind envisaged does not guarantee that reference is preserved.³¹

Dummett draws this conclusion about Kripke's remarks on naming:

We are left with this: that a name refers to an object if there exists a chain of communication, stretching back to the introduction of the name as standing for that object, at each stage of which there was a successful intention to preserve its reference. This proposition is indisputably true; but hardly illuminating.³²

But even if there is no causal theory of meaning or reference, it does not follow that Kripke has not succeeded in drawing attention to one feature of what it is for a name to refer that is an essential element of any acceptable theory. Ought we then to revise (3'C) to take account of this causal element? The case for the affirmative is well summed up by Evans:

There is something absurd in supposing that the intended referent of some perfectly ordinary use of a name by a speaker could be some item utterly isolated (causally) from the user's community and culture simply in virtue of the fact that it fits better than anything else the cluster of descriptions he associates with the name.³³

At first this argument does seem to carry some weight. In the context of a theory like (3'C), the sort of situation which Evans

has in mind would be something like the following: there is a linguistic community C in which a certain cluster of descriptions is associated with the name 'N'. Suppose that the object which best satisfies the cluster is a₁, but that a₁ is causally isolated from C and that some other object (or objects) is (are) causally responsible for those beliefs underlying the cluster. Which object is denoted by 'N'? The answer suggested by (3'C) is a₁: although members of C might intend to refer to another object, they always succeed in referring to that object which best fits what is believed about N in C.

On further reflection, however, it can be shown that the point which this argument attempts to make is already encompassed by (3'C). The difficulty with the supposed situation, construed as an objection to a cluster theory, is in understanding how the object which best satisfies a cluster of descriptions could fail to be the one which is causally responsible for those beliefs underlying the cluster. Returning to the 'Archimedes' example, how could it be that a person who satisfied a suitable majority of the descriptions 'discovered the Archimedean principle, lived most of his life in Syracuse, ran through the streets shouting 'Eureka!', and invented the Archimidean water-screw' might turn out to be causally isolated from those very deeds which are described in the cluster? The situation is palpably unreal, and this marks something of importance about descriptive theories. What a name refers to depends on what is known and believed about the bearer of the name, but beliefs and items of knowledge themselves have a causal origin. As Evans says, "the important causal relation lies between [the denoted] item's states and doings and the speaker's body of information — not between the

item's being dubbed with a name and the speaker's contemporary use of it."³⁴ But explicit recognition of this causal relation is not something that has to be added to descriptive theories, it is already contained in them. Hence, the descriptive theorist's understanding of what it is for a proper name to refer to an object does not require supplementation of this kind.

Despite the criticisms brought against the causal account, it must be admitted that there are situations in which we can imagine causal origin playing an important part in discovering what the referent of a name is. These are where there is some dispute over who or what is being talked about. Some of them might be resolved by making use of the causal relation which Kripke emphasises. Consider, for example, the following case. Erigena was a ninth century philosopher also known as John the Scot. The little that is known about him may have been true of various people, each of them having done some of the deeds attributed to Erigena. What would seem to count as a crucial factor in deciding Erigena's identity, however, would be the discovery that one of those regarded as possible bearers of the name had in fact been named 'Erigena'. The causal relation between the introduction of a name and later use of it is of some importance when those descriptions associated with the later use turn out not to be uniquely true of something. In such cases the later users would presumably accept, as having particular relevance to deciding the issue, evidence as to which of the various possible bearers of the name was so christened. This fact is readily accounted for by a cluster theory — one of the cluster of descriptions believed by users of the name 'N' to be true of a₁, if a₁ is to be counted as the bearer of the name, is 'was christened 'N''.

Which is not to say, of course, that being christened with a name is necessary in order for something to count as the referent of the name — like all the rest of the descriptions in the cluster, it is defeasible.

What lies behind this example is the point that, for a name to be introduced into a language as the name of some object, there has to be a causal relation between that object and those who introduce the name. The object might not be present at the time, but in that case the object must be causally responsible for those beliefs underlying the descriptions which are associated with the object and which are therefore used to fix the referent of the name. As regards this latter type of case, numerous instances can be imagined where the causal origin of information is of particular relevance in discovering who or what is being referred to. By way of illustration, suppose that a manuscript is found which details certain events that took place in a monastery near Florence during the twelfth century. One entry, from around 1130, refers to a visiting philosopher and theologian, but it is not clear from the script whether it was Peter Abelard or Peter Lombard. Unfortunately the remarks made concerning this person are insufficient for us to decide which of the Peters it was. Upon further investigation, however, it is revealed that at the time the entry was written Peter Lombard was studying in nearby Bologna, whereas Peter Abelard was lecturing in Paris. It would therefore seem that the latter could not have been the one whom the author of the manuscript was referring to.

In general we can say that, in those cases where there is disagreement over who or what the referent of a proper name is, one factor which may resolve the disagreement is which of the possible

bearers of the name, i.e., which of those objects satisfying a suitable majority of the descriptions associated with the name, the users of the name could have been causally related to. The point might be expressed in terms of a condition: for a proper name ' α ' to be used by members of a linguistic community C to refer to a, it is necessary that there be some causal connection between a and members of C, namely that between a and what is believed about the bearer of ' α ' in C. It is simply not possible that the members of C should be talking about something causally isolated from them. For this reason the condition might also be seen as an a priori constraint on possible attributions of referents to names.

What these remarks go to show is that although various features pertaining to what might loosely be termed the causal account of naming do not add anything to the cluster theory of naming, they are of some importance when it comes to discovering what a particular name refers to. That is, they are pertinent not to answering question (3'), but to answering the proper name version of question (2) — the question

- (2') How can we discover what the referent of a proper name is as that name is used within a linguistic community?

What (3'C) suggests by way of answer to this question is that we must consider those beliefs associated with the bearer of the name within the community. We can now add that where such a method does not enable us to decide which of several objects is the referent, the causal relations between those objects and the community can sometimes decide the issue. Also, it can be said that satisfying the causal constraint noted above is a necessary condition for an object's being counted the referent of a name. In the next section

I shall argue that these same remarks hold when we consider natural kind predicates instead of proper names.

Section (iii): A Theory of Reference for Natural Kind Predicates

Question (3) is about natural kind predicates. In section (i) of this chapter I explained how natural kind predicates could be thought of as cluster terms. I then pointed out that there was an analogous question to (3) which concerned proper names, and that proper names could also be thought of as cluster terms. In both the remainder of that section and section (ii) I discussed how this analogous question might be answered. The result was the cluster theory (3'C). If this theory does provide an adequate answer to the question of what it is for a name to refer — and we have not seen any reason for seriously doubting that it does — then it would seem that a theory just like it should answer the question of what it is for a natural kind predicate to have an extension. My proposal, then, for an answer to question (3) is

- (3A) A natural kind predicate ' ϕ ', as it is used within a linguistic community C, can be correctly applied to an object a if and only if a satisfies a suitable majority of those descriptions which could be consistently attributed to ϕ 's on the basis of what is believed within C.

Let me begin discussion of (3A) by noting that it does account for that aspect of natural kind predicates that leads us to call them cluster terms. By specifying that only a suitable majority of descriptions have to be true of an object in order for a natural kind predicate to be correctly applied to it, it is explicitly allowed that the object might not have all the properties associated with

the kind. As for what counts as a "suitable majority", this is not something that can always be specified in advance. As a matter of fact, the question does not often arise. Just as in the case of a proper name it is rare to find more than one object which satisfies a large number of the descriptions associated with the name, so too with natural kind predicates it is not often that an object satisfies only enough of those descriptions associated with the predicate that it counts as what we might call a borderline case.

This idea of an object's satisfying a suitable majority of descriptions associated with a predicate is of particular use when it comes to explaining the growth of scientific knowledge. It allows for earlier scientists being partly wrong in what they said about members of natural kinds. At the end of the previous section I indicated how (3'C) provides the basis for an answer to question (2'): we discover what a name refers to as it is used within some community by looking at what is believed in that community about whoever or whatever is the bearer of the name. (3A) likewise provides the basis for an answer to question (2): we discover what a natural kind predicate has for its extension as it is used within some community by looking at what is believed in that community about things of that kind. By not making satisfaction of every description in the cluster necessary, our investigation of what members of a previous scientific community were talking about when they used a particular natural kind term in a theory can take into account the possibility that they had some mistaken ideas about things of that kind. This point might be expressed more generally by saying that (3A) shows how a natural kind predicate might not be strictly true of anything but still have an extension.

Herein lies the core of my answer to the sort of problem that prompted, in section (ii) of Chapter 2, the four questions. There I cited some examples of where our natural inclination is to say that earlier scientists were mistaken in their theories — Bohr thought that particles have simultaneous position and momentum, Muller believed genes to be composed of proteins. In fact there are countless other examples of this, both for communities of scientists and for users of natural kind predicates generally — Ptolemy maintained that the planets revolved around the earth, Dalton thought all molecules were monatomic, whales were once thought to be fish, Mendeleeff believed that chemical elements were fundamental and irreducible, according to Aristotle the brain's function is to regulate the temperature of the body, Maxwell held that light waves must be propagated through a material ether, and so on. These descriptions themselves incline us to say that Bohr was referring to atomic particles, Muller to genes, Ptolemy to the planets, etc. The theory (3A) provides the basis of a justification for using such descriptions.

It is, however, only the core of my answer. It remains to be shown how we are to decide what beliefs earlier users had about the things to which natural kind predicates could be correctly applied. Quine's argument for the indeterminacy of translation is relevant here.³⁵ Quine has argued that, in attempting to translate from one language to another, we will be able to give logically incompatible translations which are invariant with respect to all dispositions to assent and dissent on the part of speakers of the language being translated. If we apply this to the problem of translating the languages of earlier scientists, the conclusion it suggests is that

it might be indeterminate what they were referring to. Quine's argument will be the substance of the next chapter. If Quine were right, it could well turn out that any explanation of what it is for a natural kind predicate to be correctly applied to something will be of little use when it comes to discovering which objects belong to the extension of a given natural kind predicate. Not surprisingly, I shall argue that Quine is not right, although I shall also suggest that translation is not always determinate. In his purview of translation Quine overlooks certain physical facts such as those which can be obtained by imposing a causal condition for natural kind predicates like that given for proper names. Such a condition would be: in order for a natural kind predicate ' ϕ ' to be used by members of a linguistic community C to describe ψ 's, it is necessary that there be some causal connection between ψ 's and members of C, namely that between ψ 's and what is believed about ϕ 's in C. But before getting on to this we must see if (3A) does adequately answer question (3).

As was shown to be the case with the theory of proper names (3'C), (3A) is a theory both of how the extension of a natural kind predicate is fixed and of what the sense of a natural kind predicate is. They are both functions of what is consistently believed within the community in which the natural kind predicate is used. Within the linguistic community of modern physicists, for example, the term 'electron' has as its extension just those things which satisfy a suitable majority of what they now believe to be true of electrons. Future physicists may dispute whether some of these beliefs are true of electrons; they might even accord the beliefs different weights and so arrive at a different "suitable majority"; but none of this

would jeopardize the claim they may wish to make that they have developed a better theory of electrons.

The sense of a natural kind predicate is given by those descriptions which are associated with the predicate as it is used within a community. As with a proper name, this sense is not to be thought of as given by those descriptions which some member of the community associates with the term. Sense is cognitive and hence communicable, but of course a member of a community may not fully understand a term, and what this means is that he has only a partial, perhaps imperfect, grasp of its sense.

This situation becomes more complex with terms like 'gold' which are used in everyday speech but also have a use in more technical contexts. Most of us, most of the time successfully use the term 'gold', but we would not be able to distinguish between a gold ring and a carefully prepared alloy one which contained no gold. As Putnam has pointed out, it would be absurd to say, in the light of this, that most of us just do not know what the meaning of the term is.³⁶ We need not conclude from the example, though, as Putnam does, that the extension of the term 'gold' does not depend on what is known, and therefore believed, about gold. We might follow Dummett in holding that we do fully understand the term even though we have not fully grasped its sense.³⁷ One consequence of this view is that, if the meaning of a term is just what a person knows when they understand the term fully, then a wedge is driven between sense and meaning, for as Dummett concludes, "The meaning of the word 'gold', as a word of the English language, is fully conveyed neither by a description of the criteria employed by the experts nor by a description of those used by ordinary speakers; it involves both,

and a grasp of the relationship between them."³⁸ To avoid this consequence we might stick to the simpler view that an ordinary speaker, because he has not fully grasped the sense of the term 'gold', does not fully understand it and so only knows a part of its meaning. On neither alternative, however, is one forced to give up the view that the extension of a natural kind term depends on its sense, and that sense has to do with knowledge; the only difference is that, according to Dummett, part of this knowledge consists in grasping how everyday uses of a term like 'gold' are related to its more technical uses.

Other examples designed to cast doubt on the cluster theory's account of the senses of natural kind predicates are offered by both Kripke and Putnam.³⁹ Having noted how definitions of natural kind words mention clusters of properties, Kripke attempts to show that "possession of most of these properties need not be a necessary condition for membership in the kind, nor need it be a sufficient condition."⁴⁰ Despite the way Kripke phrases this claim, it is clear that he intends his examples to be conclusive against any form of descriptive theory, irrespective of whether it is in terms of most properties or a suitable majority of properties. What they would have to show, then, to be conclusive against (3A) is that given how, say, the natural kind term 'tiger' is used within a particular linguistic community, it might turn out both that tigers should fail to possess a suitable majority of those descriptions consistently associated with them within the community, and that something could satisfy a suitable majority and yet not be a tiger.

Kripke's purported counter-example to the sufficiency condition requires us to imagine an animal being discovered which had all the

external properties mentioned in the cluster associated with the word 'tiger', but a completely different internal structure. Of course, Kripke is quite right to say that we would not count such an animal as a tiger. As it stands, though, this example does not even begin to cast doubt on a cluster theory phrased in terms of a suitable majority of descriptions. Ordinarily we identify tigers by means of properties relating to external appearance, like 'having four legs and a tawny yellow coat with transverse black stripes'. But this is not to say that we accept these properties as definitive of the species 'tiger', or that, if it came to precise definition, we should attach more weight to them than to properties relating to internal structure. Moreover, if, say, 'reptilean tigers' started appearing, properties which enabled us to identify 'mammalian tigers' would no doubt gain in importance.

The case which has to be argued in order to controvert the sufficiency condition of the cluster theory I have developed would be that an object could have a suitable majority of those properties attributable to ϕ 's on the basis of what is believed within a linguistic community C, and yet not be something to which the predicate ' ϕ ', given how it is used within C, could be correctly applied. Revising Kripke's tiger example, what would have to be maintained is that something could have all the properties that, e.g., contemporary zoologists (the experts to whom we defer) accept as definitive of membership of the species, and yet not be a tiger. It is difficult to see that any sense can be made of such a suggestion. There might come a time when, as a result of further zoological investigation, the species is defined differently; this happened with whales. Similarly, the properties now regarded as

definitive were not so regarded by earlier zoologists. But in these cases we are no longer considering the use of the natural kind predicate 'tiger' within our own linguistic community.

The situation here is just like one we encountered in connection with proper names. The referent of the name 'Archimedes', as that name is used within a community C, is that person who satisfies a suitable majority of those descriptions believed by members of C to be true of Archimedes, even if it should later turn out, as a result of subsequent investigation, that one Anaximedes is so identified. Later use of both names would no doubt reflect this discovery, with a resultant change in sense of the names. But in so far as we are considering who members of C succeeded in referring to when they talked about Archimedes, we must say it was Anaximedes.

Dummett has suggested a way of strengthening Kripke's example:

Suppose that there are on Mars creatures exactly like tigers, both superficially and in respect of internal structure. Then I think that they would still not be tigers (though doubtless they would be called 'Martian tigers'), because they would not be sprung from the same stock as real tigers, i.e., Earth tigers. A difference of internal structure serves to show that a creature is not a tiger by showing that it does not share a common descent with real tigers. For the same reason white ants are not really ants. It is a part of the meaning of a word like 'tiger' or 'ant' that it applies to an animal in virtue of its membership in a breed or family ('species' is of course too specific a term), i.e., a group connected by descent.⁴¹

Perhaps this version gains some initial plausibility from the fact that the environment on Mars is so different from that on Earth — we cannot help but think that any animal or plant found there must be of a quite different kind from any found here. But even if we make allowance for this, Dummett's example is not convincing. The

applicability of words like 'tiger' and 'ant' ultimately depends on our concept of a species. Now 'species' is a term like 'acid' in that it has a common use as well as a more specialized one. We sometimes use it to refer to a group of individuals which share a distinctive, common property, thus following its usage by 17th and 18th century zoologists and biologists. In modern biological science, though, a specimen is counted a member of some species if it can be mated or inter-bred with recognized members of the species so as to produce fertile offspring. This is the reason why white ants are not really ants, for although they might resemble them superficially they cannot form a breeding colony together. The crucial test for whether or not an animal found on Mars was a tiger, i.e., a member of the species Felis tigris, would therefore be whether or not it could mate with tigers here on Earth so as to produce fertile offspring. If Dummett's description of the alien creatures is strictly adhered to, then, it would seem certain that they could, and this would provide overwhelming evidence in favour of the conclusion that there were tigers on Mars. (Their presence there would then, for a time anyway, be a scientific anomaly.) If, on the other hand, the creatures did not mate with Earth tigers, then this would be taken as indicative of a physical difference which would tell against their being classed as members of the species Felis tigris.

All of this, however, takes us far beyond what Kripke has to say. His argument against possession of a suitable majority of properties being necessary for membership of a natural kind is even more sketchy than that against its being sufficient. He suggests it might turn out that, as a result of optical illusions or other errors, tigers actually have none of those properties characteristically

associated with them. Now of course some things might not possess such properties, and in fact most things do not possess them — only tigers do; but how could it possibly turn out that tigers might not possess them? The point to be made here is analogous to the one made in the previous section against Kripke's claim that Aristotle might not have done any of those things commonly attributed to him. It is tigers we are considering, and tigers just are those things which characteristically have four legs, a tawny yellow coat, and so on. The only sense which we can attach to the term 'tiger' — the only sense it has for us — depends on how we use it, and this reveals what we believe about tigers.

Putnam's purported counter-examples to cluster theories are more elaborate. As in the case of the 'gold' example already mentioned, Putnam's aim is to show that one cannot hold both that "the meaning of a speaker's words does not extend beyond what he knows and believes," and that "meaning determines extension."⁴² Accepting (3A) does not commit one to accepting the first of these theses; in fact, the reason we rejected it was precisely so as to take account of what Putnam calls the "principle of the division of linguistic labour." Consequently, we need not look at those examples which Putnam adduces in support of his principle and against the thesis.⁴³

Putnam's other reason for rejecting the conjunction of both theses has to do with what he calls "the contribution of the environment" to determining the extension of the terms we use. The example he presents as an illustration of this would, if Putnam's interpretation of it were correct, tell against a cluster theory like (3A). It goes like this: suppose that, on a distant planet very

similar to ours called 'Twin Earth', the colourless, tasteless liquid that comes down in rain, fills the oceans and lakes, etc., is not composed of H_2O molecules but of XYZ molecules. Furthermore, the Twin-Earthians use the term 'water' to refer to the said substance. Suppose also that we are back in 1750 and, because chemistry is underdeveloped, neither Earthians nor Twin-Earthians know the chemical structure of what they respectively call water. Thus, an Earthian and his identical twin on Twin Earth can associate exactly the same beliefs and items of knowledge with the term 'water' and its extension. Yet, says Putnam, the extension of 'water' as used by Earthians is H_2O , whereas its extension as used by Twin-Earthians is XYZ. This suggests a dilemma, either horn of which seems to lead to Putnam's desired conclusion. If we say that, because Earthians and Twin-Earthians know and believe the same things about water, 'water' means the same for both, then, since the extension is different on the different planets, there must be something more to determining extension than looking at what is known and believed. If, on the other hand, we say that because the extension is different on the different planets 'water' means something different, then this difference has to be explained in terms other than those which rely solely on what is known and believed. Therefore, either meaning does not, by itself, determine extension, or the meaning of a word extends beyond what is known and believed about what it can be correctly applied to.⁴⁴

Obviously the crucial question to be asked about the example is whether Putnam is justified in assuming that the extension of the term 'water' is different on the two planets. We can admit that after Earthians and Twin-Earthians have become chemically

sophisticated the term will have a different extension. But then, with the help of the principle of the division of linguistic labour, it is clear that we could describe the situation as one in which the meaning of the term will be different on the two planets. Before such enlightenment, however, Earthians and Twin-Earthians would agree, ex hypothesi, in their every application of the term 'water'. So it would seem incumbent upon us to conclude that when nobody knew what the chemical nature of the stuff called 'water' was, the extension of the term was the same on both planets. Once again we are reminded of the fact that the extension a term has is relative to its use within a given linguistic community.

Attention has been drawn by Zemach to an actual case very much like the one Putnam hypothesises — the discovery of isotopes of water.⁴⁵ Given the various forms of hydrogen and oxygen, together with their various combinations, it seems that we can now say that there are eighteen different kinds of water. Another actual case would be the discovery that chlorine has two common isotopes. These cases differ from Putnam's because the chlorine isotopes, and some of the water isotopes, occur together naturally and are not divided between planets. As a result it seems that, since it always was correct to apply the natural kind terms 'water' and 'chlorine' to samples containing the different kinds before isotopes were discovered, it always will be correct so to apply them.

Such examples are not likely to cut much ice (of any kind!) with Putnam. According to him natural kind predicates like 'water' — supposing it to be one — are "indexical", i.e., their extension is determined in virtue of some equivalence relation which makes use of an indexical word. In the case of 'water', Putnam suggests "the

'same liquid' relation to our water."⁴⁶ This relation is, he says, a theoretical one whose discovery is to be made through scientific investigation of whatever it is that we now identify as water. Where the natural kind predicate is a substance term the relation will typically be a structural one — chemical formula in the case of 'water', atomic number in the case of 'gold'.

This notion of indexicality is one which Putnam is keen to link with Kripke's notion of a 'rigid designator'.⁴⁷ Kripke accepts that both proper names and natural kind predicates are rigid designators — names designate the same thing, and natural kind predicates the same kind of thing, in different possible worlds. As was noted in the previous section, Kripke fills out the idea of sameness in terms of a thing's origin and structure — people could not have failed to have the parents they do, and elements could not have failed to have the atomic numbers they do. Putnam's view about 'water' — the one he thinks his Twin Earth example supports — is a short step away: water is whatever has the same chemical formula, given the present state of our chemical knowledge, as the substance we identify as water.

Returning to the isotopes, I think that even if they had been discovered on another planet and not on Earth, Putnam would dispute that they were analogous cases to his Twin Earth one. The crucial question is: which properties are the essential ones? If 'having atomic weight x ' is one, then 'water' and 'chlorine' are not natural kind predicates, for they can be correctly applied to substances with different atomic weights. But such a conclusion goes against our ordinary usage of the terms. Even chemists, when they use the terms, use them in such a way that they can be correctly applied to

separate instances, all of which share a large number of important properties. Furthermore, 'having atomic weight x' seems not to be an essential property according to Kripke's favoured test for essentiality. 'Could something be chlorine and not have atomic weight 35?' Apparently so — quite a lot of chlorine has atomic weight 37. And what this means is that a specimen of chlorine can pass all of the chemical tests for being chlorine and yet have a different atomic weight from another specimen which has also passed all of the tests. It would seem, then, that 'having atomic weight x' is not a property regarded as essential relative to our system of chemistry, for the chemical properties of a substance are not directly related to its atomic weight.

When we look at the properties 'having the chemical formula x' and 'having atomic number x', the picture is different. It does seem that we would now say 'Something could not be water (chlorine, gold) and not have the chemical formula (atomic number) H_2O (17, 79).' According to our system of chemistry, these properties do seem to be essential. What this shows, however, is not that such statements are, in Kripke's phrase, "necessary truths in the strictest possible sense,"⁴⁸ but that they are necessary relative to our system of chemistry or, more broadly, to our way of looking at the world.

Zemach points out that there have been radical changes in what science considers to be of the essence of things.⁴⁹ The very presentation which Putnam gives of his Twin Earth example goes to support this view. Despite the quotation just given, Kripke too seems to realize this, "Any world in which we imagine a substance which does not have these properties is a world in which we imagine a substance which is not gold, provided these properties form the

basis of what substance is."⁵⁰ If we gave up atomic theory, perhaps in favour of sub-atomic theory or of some non-corpuscular theory, then no doubt 'having atomic number x' would cease to be of central importance. Likewise, before 1750 there was nothing like the modern notion of an element, and so nothing could be said for distinguishing substances according to chemical formula. Hence the terms 'water', 'gold' and 'chlorine' did not, prior to the introduction of the theories and techniques of chemical analysis, have meanings which determined the course to be followed. Once this is realized, Putnam's Twin Earth example, far from showing that theses like the two he begins by distinguishing cannot be held, actually goes to support them!

Returning to Dummett's examples concerning the terms 'tiger' and 'ant', we can make a similar remark about membership of a species. The current state of the biological sciences suggests that fertile inter-breeding is an essential property of a member of a species, i.e., that it is necessary that an animal or plant, if it is to be correctly described as belonging to a species 'S', is able to inter-breed or mate with recognized members of 'S' so as to produce fertile offspring. But if we gave up the Darwinian tradition, or maybe just augmented it with a strong Lamarckian component, then this property may no longer be regarded as essential. Our understanding of what a species is may change, and with it the meaning of the term 'species'. One difference between species terms and substance terms, though, is that we seem far more reluctant to deny that an infertile animal or plant is a member of a species than we do to deny that a specimen of some substance with an unusual structure is a substance of the usual kind. Perhaps this marks a difference between the

biological and the physical sciences.

However matters stand with these "relatively essential", or "epistemologically essential", properties, accepting that there are such does not imply that there is a rival theory of reference to (3A) or that (3A) is in some way incomplete. (3A) explicitly accounts for Putnam's observations about the importance of scientific theory in fixing the extensions of natural kind predicates by dealing with suitable majorities of properties and with what is believed within the community in which a predicate is used. 'Suitable' allows us to attach more weight to some descriptions, and these might well be those suggested by the theories scientists accept. More generally, all of those descriptions which can be associated with a natural kind predicate, and hence those — if any — regarded as mentioning essential properties, will eventually be decided on the basis of what is believed within a community.

The indexical element of natural kind predicates that Putnam's theory attempts to characterize means that the essential properties are those of what is identified within a community as belonging to the kind. Because of this use of the notion of 'identification', Putnam's theory cannot be regarded as an account of what it is for a predicate to be correctly applied to an object or instance of some kind. As I have interpreted Putnam, he is drawing attention to certain contingent facts about our use of natural kind predicates, not giving a theory of what it is for a natural kind predicate to be correctly applied to an object.

One way to avoid mentioning what is identified within a community as belonging to a kind, while at the same time retaining an indexical element, would be to hold that the relation of essential

similarity holds between whatever objects are, in the Kripkean sense, causal ancestors of the present use of a natural kind predicate, and those objects which have the same essential properties. Such an account, however, would also fail to be a theory of reference for natural kinds. One reason for this is the same as the one given for why Kripke's account of naming failed to be a theory of reference for proper names: elucidation of the notion of a causal ancestor necessitates mention of an intention to refer to the same kind of thing. The appropriate causal ancestor of our use of the term 'electron' would clearly not be Stoney's, for he did not use 'electron' to refer to any kind of particle at all! To take another example, early users of the term 'gold' could not distinguish real gold from 'fool's gold', i.e., from iron pyrites, so again the appropriate causal ancestor has to be traced using some other theory. The extension of natural kind predicates can change, just as the reference of proper names can. A second reason why such an account will not do as a theory of reference for natural kinds is that, where theoretical properties are involved, a change in theory means a change in those properties regarded as essential. A cluster theory, though, can explain both of these features.

CHAPTER 6 INTERPRETING PREVIOUS SCIENTIFIC THEORIES

Section (i): Indeterminacy of Translation

Let me begin by repeating the answer I gave in the previous chapter to the question of what it is for a natural kind predicate to have an extension:

- (3A) A natural kind predicate ' ϕ ', as it is used within a linguistic community C, can be correctly applied to an object a if and only if a satisfies a suitable majority of those descriptions which could be consistently attributed to ϕ 's on the basis of what is believed within C.

What implications does this theory have for the question of how we can discover which objects belong to the extension of a particular natural kind predicate as that predicate is used within a given linguistic community? It clearly implies that we are to discover this through an investigation of what descriptions could be consistently attributed to ϕ 's on the basis of what is believed about ϕ 's within C. The first task of such an epistemological investigation, then, will be to find out what members of the given linguistic community believe, or believed, about ϕ 's.

What information will be available to one engaged on such a task? Apparently this will depend on the nature of the linguistic community whose language he is investigating. In discussing the growth of scientific knowledge, the communities will frequently be ones like '17th century French chemists', 'late 19th century physicists' and '19th century Mendelian geneticists', i.e., communities prior to our own. In such cases the information will

generally be limited to what can be deduced from physical objects which have endured since that time — books, articles, scientific instruments, and so on. For communities not so distant there might also be memories passed on — the "oral tradition" so to speak. The problem faced by the investigator is that of arriving at a characterization of the intentional states — primarily the belief states — of earlier theorists given such limited information.

One way of seeing this attempt to understand previous scientific theories is as a special case of translating from a foreign language. Rather than languages, however, the investigator will adopt as basic previous scientific theories. Beginning with established translations of relatively observational terms, he proceeds to translate the more theoretical ones until he achieves a coherent translation of statements about ϕ 's, on the basis of which he is able to impute beliefs about ϕ 's to the earlier theorists. What is assumed throughout is that the intentional states are there to be investigated. The realist views them as forming an objective subject matter, albeit one whose investigation poses difficult methodological problems.

In this chapter I wish to consider an argument of Quine's which aims to show that, given the methodological problems facing a translator, not only is translation underdetermined by all possible evidence, but that there is no fact of the matter for the translator to be right or wrong about. Translation is, to use Quine's expression, indeterminate. The conclusion he draws is that we must adopt a non-realist attitude to intentional states. If he is right, theories of the kind encapsulated in (3A) must be rejected.

What does Quine mean when he says that translation is indeterminate? We recall from Chapter 3 that, according to Quine, a speaker's knowledge of a language — the meanings he attaches to his words — is manifested primarily in his dispositions to assent to and dissent from sentences. If we wish to investigate this knowledge we must therefore attach meanings to his words in a way that accords with his verbal dispositions. The result is an interpretation of his words that is expressive of an intelligible belief set. Given this picture of verbal behaviour, Quine says:

manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of the one language, sentences of the other language which stand to each other in no plausible sort of equivalence however loose.¹

As this formulation of the thesis suggests, Quine adopts the idiom of radical translation, i.e., translation from a completely alien language. But as with the gavagai example, radical translation is really a strategy for establishing features about our own home language. The translator is then construed as one who is concerned to understand other speakers of a language with which he is familiar. As a fellow speaker he appears to be in a privileged position, for he can resort to homophonic translation and other aids. Quine maintains though that these are merely "regulative maxims" which help to settle the question of which translation to adopt, not which translation captures the real meanings. The methodological problems are claimed to hold equally for the radical translator and

the fellow speaker of English. The thesis may be re-stated as the infinite totality of sentences of any given speaker's language can be so permuted, or mapped onto itself, that (a) the totality of the speaker's dispositions to verbal behaviour remains invariant, and yet (b) the mapping is no mere correlation of sentences with equivalent sentences, in any plausible sense of equivalence however loose.²

So far it might seem that Quine is trying to establish an epistemological point to the effect that all possible observations of verbal behaviour fail to determine a unique interpretation of the sentences of a language. But what makes Quine's whole argument so contentious is that he wishes to make the much stronger ontological or metaphysical point that there is simply no question of one of the interpretations being true and the rest false:

"The point is not that we cannot be sure whether the analytical hypothesis is right, but that there is not even, as there was in the case of [the occasion sentence] 'Gavagai', an objective matter to be right or wrong about."³ What argument does Quine have to support this conclusion?

What Quine calls "the real ground" of the doctrine that translation is indeterminate is an inference from the underdetermination of scientific theory by all possible observations. He describes this as follows:

If our physical theory can vary though all possible observations be fixed, then our translation of [a foreigner's] physical theory can vary though our translations of all possible observation reports on his part be fixed. Our translation of his observation sentences no more fixes our translation of his physical theory than our own possible observations fix our own physical theory.⁴

The crucial point, however, is his insistence that "The indeterminacy of translation is not just an instance of the empirically underdetermined character of physics."⁵ Thus, in an often quoted passage from his reply to Chomsky, Quine says:

Though linguistics is of course a part of the theory of nature, the indeterminacy of translation is not just inherited as a special case of the underdetermination of our theory of nature. It is parallel but additional.. Thus, adopt for now my fully realistic attitude toward electrons and muons and curved space-time, thus falling in with the current theory of the world despite knowing that it is in principle methodologically underdetermined. Consider, from this realistic point of view, the totality of truths of nature, known and unknown, observable and unobservable, past and future. The point about indeterminacy of translation is that it withstands even all this truth, the whole truth about nature. This is what I mean by saying that, where indeterminacy of translation applies, there is no real question of right choice; there is no fact of the matter even to within the acknowledged underdetermination of a theory of nature.⁶

This stage of the argument will be discussed at length in the third section. Before we come to it, let me make some further clarificatory points.

Despite his calling the argument from underdetermination of theory by evidence "the real ground" of the doctrine of indeterminacy, Quine nevertheless holds that the gavagai example suggests another ground. Since 'Gavagai!' is an observational sentence, it has a stimulus synonymous translation and is itself "an example only of the inscrutability of terms, not of the indeterminacy of translation of sentences."⁷ To see where it does support indeterminacy, we need

to imagine "that some lengthy non-observational sentences containing gavagai could be found which would go into English in materially different ways according as gavagai was equated with one or another of the terms 'rabbit', 'rabbit stage', etc."⁸ Yet here we must bear in mind the conclusion of Chapter 3 — that even if we accept Quine's stricture that only behavioural facts are relevant in translation, we can establish a unique translation for a predicate like gavagai. It will then be irrelevant whether we are talking about its use in observational or non-observational sentences. For these reasons, I shall confine my attention to the under-determination of theory by evidence when discussing what argument Quine gives in support of his doctrine of indeterminacy of translation.

Having said this, the relation between indeterminacy of translation of sentences and inscrutability of reference of terms still remains unclear. Quine says "the inscrutability of terms need not always bring indeterminacy of sentence translation in its train."⁹ Apparently then inscrutability does not imply indeterminacy. From this it follows that determinacy of translation does not imply 'scrutability' of reference, i.e., we might be able to determinately translate some foreign sentence without being able to 'scrute' the reference of its terms. This makes sense of what Quine says about his examples of the Japanese classifiers and deferred ostension.¹⁰

In the context of this thesis, the important question is whether the converse implication holds, i.e., whether scrutability of reference implies determinacy of translation. One of my aims is to establish what the terms of earlier scientific theories had as their extensions. If scrutability implies determinacy, indeterminacy

implies inscrutability, and consequently our inability to determinately translate sentences would mean that we cannot be assured of establishing what the extensions of their terms are. Moreover, it would seem that indeterminacy often pervades our interpretations of previous scientific theories. We are, e.g., unsure as to precisely what was believed by Mendel about "formative elements", even though we interpret them as beliefs about genes. If the implication holds, and if there is no fact of the matter about what Mendel believed, such an interpretation is seriously questioned. The situation will be like the one faced by the realist in the wake of the incommensurability thesis. Given the incommensurability of two competing or successive scientific theories, it was claimed that there was no fact of the matter as to whether they are about the same things. This was held to be so on the grounds that there would be too few "translations" — in fact none — between the theories. Should inscrutability of reference be held to follow from indeterminacy of translation, the realist will face the same conclusion, although this time the reason will be that there are too many translations all compatible with the behavioural evidence.

If, on the other hand, scrutability does not imply determinacy, then, by similar reasoning to that appealed to in the last paragraph but one, indeterminacy does not imply inscrutability. It would then be possible for the realist to concede that translation might not be fully determinate without thereby jeopardizing his claims to have established what the extensions of terms from previous theories were. And indeed there does seem to be good intuitive support for this view.

Simply because a radical translator can tell, by using procedures like those outlined in Chapter 3, that members of an alien tribe are referring to rabbits, and not rabbit stages, etc., when they use the term gavagai, it does not follow that he will thereby be able to interpret their whole "theory" of rabbits. It might be the case, e.g., that they have beliefs of a religious kind about rabbits, beliefs which remain indeterminate given their stimulus responses. Likewise someone might be able to show that Dalton had a theory about atoms without being able to state the totality of his beliefs about them — if he can establish a suitable majority, that is enough. This is clearly not Quine's view. I suggest it only as one line of argument which might be followed if translation could not be shown to be indeterminate in a way additional to the underdetermination of physics. We shall return to it later.

For the moment, let us concentrate on the first two stages of the argument Quine calls "the real ground". These are: (i) assume that scientific theory — "the theory of nature" — is underdetermined by all possible observations, then (ii) it follows that translation is underdetermined by all possible observations. A first query might be raised over the phrase "all possible observations": how, one might ask, could we ever be in a position to rule out all future evidence? But there is a weaker formulation of which Quine can avail himself. There is no reason to think that at any stage of scientific enquiry the totality of observations we have made will force us to accept just one theory; alternatives will always be available.

Increased recognition of the work of Duhem has resulted in general acceptance of this view. Indeed, in most quarters it has come to be regarded as something of a platitude. Now although I think there are interesting questions to be raised about particular purported instances of underdetermination of scientific theory, it does seem that there are numerous examples which are correctly interpreted in this way.¹¹ Hence I shall follow the vast majority of Quinean commentators in conceding this first assumption.

Turning now to translation, we need to begin by examining the methodological situation of the radical translator. His aim is to define a function which maps sentences of the alien language onto sentences of the translator's home language. The basic data, as already noted, are the aliens' dispositions to assent to and dissent from sentences, it being assumed that native assent and dissent can be interpreted as such by the translator. In addition to this evidence, Quine thinks there are certain constraints which the translator is methodologically justified in imposing on the function. Observation sentences must be mapped onto sentences of the home language generally assented to in the same circumstances. Truth-functional logic should be imputed to the aliens. Finally, stimulus-analytic (-contradictory) sentences of the alien language should be translated by stimulus-analytic (-contradictory) sentences of the home language. The resultant possible functions Quine calls analytical hypotheses.

Two problems affecting this stage of the argument may be distinguished. The prior one is whether Quine has described the methodological situation correctly. The second is whether, given a

correct description, translation is underdetermined by observation. Discussion of the first centres on the issue of what the correct constraints on translation are. Some critics have suggested that Quine is unjustified in imposing the three that he does. Thus Hookway notes, with respect to the third, that the body of stimulus-analytic sentences changes over time, with the result that a sentence like 'The sun circles the earth', which may have been stimulus-analytic in the middle ages, is now more likely to be regarded as stimulus-contradictory.¹² The more frequent criticism of Quine though is that he ignores other acceptable constraints. I shall consider this criticism in detail in section (ii). The problem of how determinate translation is, given a correct description of the methodological situation, will be discussed in the third section.

Section (ii): Radical Translation vs. Radical Interpretation

Quine has frequently drawn attention to the close connections between belief and meaning. In order to determine what sentences of an alien language mean we must decide what beliefs they express. But how can we begin to attribute beliefs without some means of interpreting the language? A start has to be made somewhere, yet any particular point would seem to exceed what is justified by the behavioural facts.

In Chapter 3, I noted one reply to this line of argument which turned on the contrast between a theory of translation and a theory of interpretation. Attention was drawn to the fact that

beliefs interact with desires to determine actions. This suggests a way in which the data used to establish translation can be enlarged. In translation the "variables" are belief and meaning; in the interpretation of action they are belief and desire. Perhaps it would be possible to "play off" verbal behaviour and action in such a way as to fix on belief.

Unfortunately it is difficult to see how such a reply could offer real aid, in the form of further constraints, to the radical translator. It is as though instead of having one equation with two variables he now has two equations with three variables; for a unique solution he needs one more equation with no more variables. Of course, he could impose certain a priori restrictions on the plausibility and complexity of the desires ascribable, but any such assumptions about intentional states would seem to beg the question against Quine's argument. For this reason it seems to me that the approaches of Lewis¹³ and Grandy¹⁴, whereby aliens' beliefs, desires, and world pictures are simply assumed to be as similar to our own as possible, require further clarification. Moreover, as Hookway has noted, even assumptions made about the intentional states of those belonging to the same linguistic community as a translator, assumptions he might intend to impute to the aliens, stand in need of independent support: "If psychological generalizations are grounded in observations of the behaviour of those whose behaviour we understand, they cannot provide an essential empirical constraint rendering interpretation determinate."¹⁵

A more promising approach to the belief/meaning problem in radical translation is suggested, somewhat ironically, by Quine himself. In his discussion of the translation of sentential connectives, he notes that there will be cases where the translator is obliged to reconstrue the behavioural evidence. This will be so where the implied translations result in "assertions startlingly false on the face of them."¹⁶ He cites approvingly Wilson's "principle of charity": "We select as designatum that individual which will make the largest possible number of ... statements true."¹⁷ This principle is so strong that, even in translation at home, "We will construe a neighbour's word heterophonically now and again if thereby we see our way to making his message less absurd."¹⁸

Maximizing truth in this way is nothing more than maximizing agreement between speaker and interpreter over particular statements, for the truth of those statements is judged by the latter. The principle of charity, then, is a constraint on translation, one which rests on the idea that speaker and interpreter are to be assumed as sharing certain beliefs. Davidson has recently argued that this principle is the key to overcoming systematic indeterminacy.¹⁹ His claim is that a theory of radical interpretation can, with the aid of the principle, be constructed in a way formally analogous to Tarski's definition of truth. Where he differs from Quine over the principle is in holding that it has to be applied "across-the-board",²⁰ not just in connection with the identification of purely sentential connectives. The remainder of this section will be spent examining these views of Davidson's.

Following Quine, Davidson maintains that the crucial notion for a theory of radical interpretation is that of accepting sentences as true. Given the close interrelation of belief and meaning, the only way to then begin interpreting a speaker's words is by assuming general agreement on beliefs:

We get a first approximation to a finished theory by assigning to sentences of a speaker conditions of truth that actually obtain (in our own opinion) just when the speaker holds those sentences true. The guiding policy is to do this as far as possible, subject to considerations of simplicity, hunches about the effects of social conditioning, and of course our common sense, or scientific, knowledge of explicable error.²¹

Such a procedure "is not designed to eliminate disagreement, nor can it; its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation — some foundation — in agreement."²² Here we have the basis for a transcendental argument in support of the principle of charity:

Since charity is not an option, but a condition of having a workable theory, it is meaningless to suggest that we might fall into massive error by endorsing it. Until we have successfully established a systematic correlation of sentences held true with sentences held true, there are no mistakes to make. Charity is forced on us; whether we like it or not, if we want to understand others, we must count them right in most matters. If we can produce a theory that reconciles charity and the formal conditions for a theory, we have done all that could be done to ensure communication. Nothing more is possible, and nothing more is needed.²³

One person to have questioned this argument of Davidson's is Colin McGinn. He claims that

we may equally provide a basis for deriving the meanings of

sentences held true by uncharitably imputing false beliefs to our speaker. We simply suppose, with or without good reason, that he has made a mistake and is expressing a false belief with a correspondingly false sentence.²⁴

Davidson, however, appears to have foreseen this reply, for he says elsewhere

The methodological advice to interpret in a way that optimizes agreement should not be conceived as resting on a charitable assumption about human intelligence that might turn out to be false. If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything.²⁵

McGinn also draws attention to another argument Davidson has for charity, that before someone can be said to have a belief about something it has to be shown that they have many other true beliefs about that thing.²⁶ He rightly points out that this condition is too strong as it stands, for there do seem to be cases where we can justifiably say that a person has a belief about something without his having many beliefs about it at all. I fail to understand McGinn's further point though that

It may still be maintained ... that possession of a concept requires a certain minimum of true beliefs about members of its extension, so that there cannot be shared concepts without a measure of shared beliefs; but this falls short of what Davidson wants, because now we see that disagreement concerning an object is possible unmediated by common concepts with respect to that object.²⁷

To refashion an example of Davidson's, what convinces us that some ancients believed of the earth that it was flat is that they made

a large number of statements which we interpret as true statements concerning the earth and which we take as expressing true beliefs; statements about, e.g., the physical shape and climatic conditions of particular parts of its surface. Certainly we have good reason for thinking that they did not share our concept of the earth as a large, cool, solid body circling around a very large, hot star, but as I have emphasised all along, identity of reference is different from identity of sense, and when we talk about "common concepts" we understand the latter to be the case. If this is so then, contra McGinn, such examples do seem to support Davidson's claim that "False beliefs tend to undermine the identification of the subject-matter; to undermine, therefore, the validity of a description of the belief as being about that subject. And so, in turn, false beliefs undermine the claim that a connected belief is false."²⁸

The point here is just like the one I made in response to Kripke's purported counterexamples to the account given by the cluster theory of how the referent of a proper name or the extension of a natural kind predicate is fixed. Before we can make sense of assertions about what Aristotle might not have done, or about what properties tigers might not have possessed, we have to have some way of picking out the relevant subject-matter, and this limits the beliefs we might wish to entertain. Consequently it would seem that if we replace, in the last quotation from McGinn's article, the phrase "a certain minimum of true beliefs" with the phrase "a suitable majority of true beliefs", and understand "a suitable majority" in the way suggested in Chapter 5, we have another good argument in favour of the principle of charity. Moreover, it will be an argument supported by what was said in Chapter 5.

In the light of these arguments, considerable charity appears to be obligatory on the interpreter. The next question is, how is he to apply it across-the-board? To answer this we need to examine Davidson's theory of radical interpretation more closely. One of his main ideas is that a theory of truth for a language should enable a person to understand any declarative sentence uttered by a speaker of that language. If one knows that when a speaker utters X what he says is true if and only if p, one can grasp what he is claiming to be true, if he is asserting X. It would seem to follow from this that one would be justified in concluding that X means p. But here Davidson says that we have to seriously modify the initial idea that a theory of interpretation is to be based on a theory of truth. Tarski's Convention T demands of a theory of truth that all sentences of which a truth predicate is predicated entail others of a certain form. This gives sentences like the familiar '"Frege died in 1925" is true if and only if Frege died in 1925'. Following Davidson, let us call these T-sentences. In Tarski's theory, T-sentences are to be recognized by their syntactic form, but in radical interpretation a syntactical test is not available since it would presuppose an understanding of the language to be interpreted. As Davidson says:

the syntactical test is merely meant to formalize the relation of synonymy or translation, and this relation is taken as unproblematic in Tarski's work on truth. Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp of the concept of truth. What we require, therefore, is a way of judging the acceptability of T-sentences that is not syntactical, and makes no use of the concepts of translation, meaning or synonymy, but is such that acceptable T-sentences will in fact yield interpretations.²⁹

Such a way is afforded, according to Davidson, by the principle of charity.

If speakers of a language hold a sentence to be true under certain circumstances observed by the interpreter, then this is to be taken as prima facie evidence that the sentence is believed by the speakers to be true under those circumstances. A speaker is thereby assumed to be truthful as far as possible. In this way the interpreter can hope to establish, via numerous positive instances, generalizations like

(x)(t)(if x is a member of the German linguistic community
then (x holds 'Es schneit' true at t if and only if
it is snowing near x at t))

which in turn support T-sentences like

'Es schneit' is true-in-German for a speaker x at time t if
and only if it is snowing near x at t.

Here we see an important difference between radical interpretation and radical translation: such reference to objective features of the world which alter in conjunction with changes in attitude towards the truth of sentences replaces Quine's notion of stimulus meaning. Notice also that the appeal to the notion of a linguistic community begs no question, for speakers can be said to belong to the same linguistic community if the same theory of interpretation works for them.

As Davidson is the first to admit, this strategy presents obvious empirical difficulties. Speakers may be wrong about whether it is snowing near them; there will be differences from speaker to speaker, and from time to time for the same speaker,

with respect to the circumstances under which a sentence is held true; and so on. Davidson replies by pointing out that a theory of interpretation will have to take account of the holistic nature of a language. This gives another (possibly unintended) sense to his remark that charity has to be applied across-the-board. T-sentences are not to be established one at a time but rather as elements of a pattern which satisfies the formal constraints of a theory of truth. The aim is to get a theory of best fit, although there will be no reason to suppose that there is just one such theory.

This raises the important question of how much indeterminacy there will be in radical interpretation. Before we can assess this we need to look more at the formal constraints. Beginning with those sentences always held true or always held false, i.e., those identified in Quinean radical translation as stimulus-analytic or stimulus-contradictory, and patterns of inference, the radical interpreter looks for the best way to fit his logic, to the extent required to get a theory satisfying Convention T, onto the alien language. Logic is here treated "as a grid to be fitted onto the language in one fell swoop."³⁰ It seems that this will be possible only if he can find, in the alien language, structures of first-order logic required by the theory for the proofs of T-sentences. So every language, if it is to be counted a language at all, will have an underlying logic identical to our own, and this will immensely limit the admissible translations. Identifying connectives, singular terms, predicates, quantifiers, and identity settles matters of logical form. Indexical sentences, whose truth-value is relative to the environment, are interpreted next, in the way outlined two paragraphs ago. Finally, the interpreter

tackles those sentences whose truth-value neither commands uniform agreement nor depends systematically on changes in the environment. The hope is that the recursion demanded by the theory will lead to their determinate interpretation.

Despite these remarks, it must be said that Davidson remains vague, perhaps even unsure, about how determinate translation will be. He comments:

There may, as Quine has pointed out in his discussions of ontological relativity, remain room for alternative ontologies, and so for alternative systems for interpreting the predicates of the object language. I believe the range of acceptable theories of truth can be reduced to the point where all acceptable theories will yield T-sentences that we can treat as giving correct interpretations, by application of further reasonable and non-question-begging constraints. But the details must be reserved for another occasion.³¹

So far the occasion appears not to have presented itself. In the next section I shall question Davidson's concession to Quine here and propose two "non-question-begging constraints", based on remarks made in previous chapters, which are pertinent to the problem of interpreting previous scientific theories. I shall also consider other objections to Davidson's strategy and finish with a recommendation about how we should view translation. Before doing so, though, let me briefly summarize the argument of the chapter up to this point.

Quine maintains that, both in radical translation and in translation of our own home language, it is possible to establish, given all the behavioural evidence, incompatible translation manuals which "stand to each other in no plausible sort of equivalence

however loose." His argument for this view depends on physical theory being underdetermined by evidence. He admits that some sentences (the observational ones), and the logical connectives, will be determinately translatable, but denies that there is any path leading from them to determinate translations of more theoretical sentences. Moreover, given the interdependence of belief and meaning, there will be no way of ascribing beliefs to speakers or meanings to their words. One is then invited to conclude that there is no fact of the matter about correct translation. If Quine is right, and if indeterminacy implies inscrutability, then severe limitations are imposed on any attempt to discover, in the way suggested in this thesis, the extensions of terms from previous scientific theories.

Davidson, broadening the scope of the enquiry from translation of language to interpretation of behaviour and extending some of Quine's earlier remarks, argues that there are certain assumptions that must be made if we are to be able to interpret a language at all. It must have quantification theory as a base, and the speakers of the language must be assumed to be right as often as plausibly possible. The latter assumption solves the problem of the interdependence of belief and meaning "by holding belief constant as far as possible while solving for meaning."³² The former assumption paves the way for modelling a theory of interpretation on a theory of truth. In conjunction with such a model, the assumptions so constrict the possible translation manuals that, far from standing to each other "in no plausible sort of equivalence however loose," the only indeterminacy is that which affects the predicates.

Section (iii): How Determinate is Translation?

Since Davidson's theory of radical interpretation is intended as a reply to a doctrine of Quine's, a good place to begin criticism of the theory would be with what Quine has to say about it. Fortunately Quine has recently published some comments on Davidson's proposals.³³ He pronounces himself in agreement with the strategy of taking the theory of truth as basic, and appears well-disposed, if empirically sceptical, about applying charity across-the-board as a means of disentangling belief and meaning. Despite this, he still foresees two areas where indeterminacy will arise.

The first of these involves inscrutability of reference. Quine agrees with Davidson that "Tarski's truth construction can't be carried through until we've decided what to count as quantification, or the equivalent referential apparatus, in the object language,"³⁴ but points out that in Word and Object he argued that such decisions depend on analytical hypotheses and so are not unique. He confesses though, as we noted in Chapter 3, that recent reflection on substitutional quantification has made him more tentative on this point, "So maybe one should be more hopeful about the near-uniqueness of the manual up to the point where the truth definition can be brought to bear."³⁵ None of this is said to impugn ontological relativity, however, since the values of the variables are not fixed by decisions about quantifiers. What is more, one can have different Tarskian truth definitions, delivering the same totality of expressions as true, yet differing in values

assigned to variables. According to Quine, then, inscrutability of reference threatens even in the theory of radical interpretation.

In Chapter 3, I discussed at length Quine's arguments concerning ontological relativity, and the conclusions drawn may now be advanced to support Davidson's position. To begin with, I gave, at the beginning of section (iii) of Chapter 3, several guidelines which could be used by the radical translator to deduce which alien expressions, if any, correspond to our singular terms. I then explained how he could apply this information, together with that suggested by some of Quine's reflections on substitutional quantification, in a way that would allow him to preserve logical form in translation. And this is exactly what Davidson requires, given his initial assumptions, in order for the first part of his proposed construction of the theory of truth for the alien language to go through. Quine rightly points out that even though we might have to assume that the alien language has quantification theory as a base Davidson has given no explanation of how to decide which alien expressions count as quantifiers. Chapter 3, however, does contain such an explanation.

The third stage of the argument in Chapter 3 made use of some arguments of Evans's which emphasised the deep connection between identity and predication. They suggested how, given the translator's newly-acquired knowledge of logical form in the alien language, it would be possible to determine whether or not the alien identity predicate is to be translated by our identity predicate. Then, in the final stage, it was shown how to resolve the quandary over translating the predicate gavagai. Furthermore, it was pointed out

that at no stage of the argument is it necessary to assume the translator has anything other than behavioural evidence with which to work. He can accept this restriction and still find nothing relative about the alien ontology. We can therefore meet Quine's first objection to Davidson's proposals, as well as ease Davidson's own conscience on the matter.

His second objection relates to the later stages of the construction of a theory of truth for an alien language. Whilst conceding that "We can settle the truth of [the T-sentence for 'Es schneit'] as well as we can settle the translation of 'Es schneit' into 'It's snowing'," Quine thinks that "when we get off to sentences remote from observationality we're going to have the problem of indeterminacy of translation."³⁶ Clearly the issue here is the underdetermination of theory by evidence. Allowing for determinate translation of what, for the sake of argument, we might agree to call observational sentences, Quine is claiming that translation of non-observational sentences still remains indeterminate — there is no fact of the matter about it.

In reply, Davidson concedes that indeterminacy will probably enter here, but he thinks the formal constraints imposed by a theory of interpretation will keep it to a minimum. The idea underlying this claim is best expressed in the following passage:

If we consider any one T-sentence, this proposal [that T-sentences should be true] requires only that if a true sentence is described as true, then its truth conditions are given by some true sentence. But when we consider the constraining need to match truth with truth throughout the language, we realize that any theory acceptable by this standard may yield, in effect, a usable translation manual running from object language to

metalanguage. The desired effect is standard in theory building: to extract a rich concept (here something reasonably close to translation) from thin little bits of evidence (here the truth values of sentences) by imposing a formal structure on enough bits.³⁷

The formal structure is determined by the fact that any adequate theory of truth based on the Tarskian paradigm has to be finitely axiomatizable and satisfy Convention T. It follows from this that the theory will be recursive. Apparently, then, Davidson's hope is that if we attend to enough "thin little bits of evidence" we shall be able to translate more and more complex sentences.

These remarks are brief and highly speculative. In the first place we have to distinguish between complex sentences and non-observational sentences. One can generate arbitrarily complex sentences using sentential connectives and observational sentences, but as we saw in Chapter 1 there is in general no reduction of the theoretical to the observational. Davidson's position needs supplementing here: how are we to get from the "bits of evidence" to the "rich concept"?

A promising approach is suggested, again somewhat ironically, by taking seriously the Duhem/Quine network model of theories and extending it, as Quine does in "Two Dogmas of Empiricism,"³⁸ to language as a whole. The picture which Quine presents there is one of language as an articulated structure which makes contact with reality only at the periphery. Such a picture is intended to express the fact that the sentences of a language are related by various inferential connections, and that our understanding of any one sentence involves our apprehension of such connections. It is

not necessary to discuss the ramifications of this view just now. I introduce it only as a way of making more vivid the point that even the least observational of sentences, i.e., those furthest from the periphery, are inferentially linked to others less far from the periphery.

We must also bear in mind another point made in Chapter 1: there is no firm distinction to be drawn between theoretical and observational terms. A sentence is commonly described as "theoretical" because one or more of its terms are thought of as theoretical. But if no sense can be made of any firm distinction between theoretical and observational terms, then presumably none can be made of a firm distinction between theoretical and observational sentences. Quine, it is true, often talks of "observation sentences", giving the impression that he thinks they constitute a distinct class, but when he is precise he recommends that one speak rather of "observational sentences" or in terms of "degrees of observability."³⁹

What these remarks suggest is that the translator will have to attend closely to the various ways in which linguistic evidence can intersect with theoretical structure. Theory is underdetermined by evidence, but any body of sentences deserving to be called a theory has observational consequences and so has links with observational sentences. The formal structure of the truth theory requires a recursive generation of T-sentences across-the-board, from the relatively observational through the whole spectrum to the relatively non-observational. Hopefully, then, the overall pattern gives rise to the rich concept.

The need to attend to the interrelations between the observational and the non-observational is a point that was briefly discussed in section (iii) of Chapter 2. We were considering how it was that Dalton's atomic theory of gases could be understood by his contemporaries, particularly by Avogadro who proposed a theory which conflicted with Dalton's. The point was made that Dalton's theory, like any other, did not suddenly appear, couched in a language quite removed from that of the physical science at the time. To put the matter in a rather Feyerabendian way, terms used in the theory had meanings partly determined by their links with other scientific terms in use at the time. Dalton talks, e.g., of the "atoms" or "ultimate particles" of a gas as being those which are irreducible given the known techniques of "chemical analysis and synthesis."⁴⁰ Experiments are then described involving these particles, and Dalton comments on the findings of Gay-Lussac and others in their work on gases. Theoretical terms are constantly used in discussing experimental consequences and in the interpretation of previous work, and hence the theory was intelligible to a wider audience.

Towards the end of the same section it was noted that this made sense of Parsons' proposal to test a "hypothesis" as to the extension of a scientific predicate against the "informational background" of the theory in which it occurs. Parsons says that it may prove necessary to defend such a hypothesis "against the actual arguments given by the author in question (or under the theory in question),"⁴¹ and this I take to be in the spirit of Davidson's remarks that conformity to alien beliefs is only a first assumption and that revision might later be called for at certain points.

In talking of earlier scientific theories we are moving away from the field of radical translation or interpretation and coming closer to home; our concern is not with present speakers of an alien language but with past speakers of a language very similar to our own. Such a move will mean that the translator will no longer be able to observe responses to verbal prompting, although qua interpreter his primary concern will still be with sentences held true and the principle of charity will still apply. This similarity might incline one to adopt a Quinean position to the effect that the only significant difference lies in the regulative maxims we are prepared to apply. I shall discuss this shortly. First I want to consider one more possible objection to Davidson's approach which draws on what was said in an earlier chapter, and then try to reach some conclusion about how determinate radical translation will be in the light of what has been discussed so far.

The objection can be put in the following way: In Chapter 4 you argued that Tarski's theory is a correspondence theory and that it presupposes a determinate relation of reference between expressions of a language and parts of the world; reference, you said, in this way underwrites truth. What you claimed the theory fails to explain is how we are to discover which object, if any, a singular term refers to, and which objects, if any, belong to the extension of a predicate. This led you to enquire into the problems of what it is for a proper name to refer or for a natural kind predicate to have an extension. Your answers involved cluster theories of reference, and these theories are phrased in terms of the beliefs that can be ascribed to users of the proper name or

predicate. Yet here you are in the following chapter endorsing Davidson's use of Tarski's theory to resolve problems of belief ascription. Is there not something circular in this?

My response to this charge is that it rests on a confusion of questions (2) and (3), the very confusion that was warned against in Chapter 2. Davidson's proposals are intended to find a way round some arguments of Quine's designed to show that there is no fact of the matter about how to translate certain sentences. The relevance of Quine's conclusion to the four questions at the heart of this thesis lies in the suggestion that we have no good reason for thinking that, in general, we shall be able to discover what the extensions of natural kind predicates occurring in previous scientific theories are. Chapter 5 contained my answers to the conceptual questions (3) and (3') concerning what it is for a name to refer or for a natural kind predicate to have an extension. The implication of (3A) — my answer to (3) — for question (2), the epistemological question of how we can discover what a particular natural kind predicate 'P' has as its extension, was said to be that we have to decide what beliefs users of 'P' had about those things to which they applied the term 'P'. If Quine were able to show that we are forced to adopt a non-realist attitude to beliefs, such a position would be moribund. But both he and Davidson recognize that some logical theory has to be imputed to alien speakers if we are to understand their language at all, and that this can only be done if we are prepared to ascribe beliefs to them in a charitable manner. In both radical translation and radical interpretation, then, it does make sense to ascribe beliefs, so (3A) is not ruled out on

a priori grounds. The main point of the argument of Chapter 4 was to show that Tarski's definition does not give a theory of reference but presupposes one. Chapter 5 explains what such a theory would look like. The problems of recursively defining T-sentences for an alien language and answering the epistemological question (2) can then be tackled simultaneously: we wonder what the extension of the Greek $\chi\rho\nu\sigma\delta\varsigma$ was; we make charitable decisions about T-sentences in which it occurs in such a way as to maximize agreement with other T-sentences. The important point is that ascriptions of belief, and hence translations resulting from particular T-sentences, are revisable — further evidence can always be accommodated. The residual problem then is of how determinate we can be in translation, and this affects our answer to question (2).

Let us see if we can reach some tentative conclusion about how determinate translation will be if the programme of radical interpretation is carried out. Since quantificational structure is presupposed in the alien language the radical interpreter will be able to overcome, as noted above, problems of logical form and ontological relativity (of the pervasive kind suggested by Quine's gavagai example anyway). Davidson is thus misguided in his reason for thinking there might still be indeterminacy at the level of predicates. Nevertheless, it is not clear from what he says that there will be no reason for thinking that there will be no indeterminacy at the predicate level. And despite the above remarks about the interrelations of the theoretical and the observational, one might still feel that Quine has given, in his second objection, good reason for thinking there will. Although there cannot be two

translations of a sentence, one, say, in subject-predicate form and the other not, we have as yet seen no conclusive reason for thinking there will not be many cases where an alien sentence s — "remote from observability" — may be translated by either of the home subject-predicate sentences p and q, even though p and q have different truth-values.

Hacking has argued that such a situation could not arise in radical interpretation.⁴² The requirement of radical interpretation is that such statements of translation must match with T-sentences,

So there will be nothing to choose between a T-sentence 's is true if and only if p', provided in one system, and 's is true if and only if q', provided in another system. Yet we had the initial overriding requirement that T-sentences are true. Since p and q may be contraries, both T-sentences cannot be true. From Davidson's standpoint, this is a reductio ad absurdum of indeterminacy.⁴³

What Hacking overlooks, though, is the empirical nature of radical interpretation. Davidson emphasizes that the theory of truth for an alien language "is tested by evidence that T-sentences are ... true."⁴⁴ Elsewhere he says:

If we treat T-sentences as verifiable, then a theory of truth shows how we can go from truth to something like meaning — enough like meaning so that if someone had a theory for a language verified only in the way I propose, he would be able to use that language in communication.⁴⁵

So on the basis of all the behavioural evidence we still might not be able to decide between incompatible theories of truth, and hence between incompatible translation manuals, despite their being equivalent to each other in point of logical form.

Davidson's theory of radical interpretation might be thought of as an abstract reply to an abstruse doctrine. If Davidson is right, Quine's description of the methodological situation of the radical translator is not so much wrong as inappropriate. Yet even if we agree that the radical interpreter is more faithful to our picture of one starting from scratch to understand an alien language, the claim that indeterminacy will be greatly reduced still looks highly speculative. Recognition of the interrelations suggested by the network model of language might help to alleviate some scepticism, but it scarcely constitutes a telling reply to Quine's remarks on the underdetermination of theory by evidence. What I propose to do now is to make some final points on this issue, concentrating on how it affects our interpretation of earlier scientific theories.

Suppose we are trying to interpret Dalton's atomic theory of gases. The primary evidence we have available, let us imagine, consists of his published writings dealing with the theory, written communications with others, some laboratory notebooks, discussions and interpretations of his work by others, and some scientific instruments he was known to have used. From his own writings we manage to deduce a body of sentences asserted by him and hence, we assume, believed by him to be true. We begin to interpret these sentences, starting with the more observational, in such a way as to make Dalton consistent, and correct as often as possible. Since we have no difficulty in recognizing and translating the sentential connectives and quantificational devices, this proceeds quite smoothly, although we will doubtless have to revise earlier translations and belief ascriptions as we go along. In particular, we must remember

that we are guided by our own scientific lights here. Our present theories constitute our "theory of nature" and so we must reckon them the standard by which we are to judge past theories.

If Quine is right in maintaining that there is no difference in principle between attempts to understand alien languages and attempts to understand any others, then there will be definite limits to how far we can go in thus understanding Dalton:

Insofar as the truth of a physical theory is underdetermined by observables, the translation of the foreigner's physical theory is underdetermined by translation of his observation sentences. If our physical theory can vary though all possible observations be fixed, then our translation of his physical theory can vary though our translations of all possible observation reports on his part be fixed. Our translation of his observation sentences no more fixes our translation of his physical theory than our own possible observations fix our own physical theory.⁴⁶

The interrelations of the theoretical and the observational suggested one reason for doubting this. In similar vein we might draw attention to the wide range of writings available as evidence. All of these will have their own situations, some involving connections with other scientific communities which will in turn be interrelated with larger linguistic communities. We might call this 'secondary evidence'. What has to be remembered is that the interpretation is based on a theory of truth for a language, and so it should not be surprising if the evidence is vast.

Are there any other reasons for doubting Quine's claim? Another is forthcoming when we reflect on the fact that we think there are, at most, only a small number of things Dalton could possibly have been talking about. Given our knowledge of the state of physical science

at the time, we can make empirically justified statements about what Dalton could have acquired knowledge of. We can form some idea, that is, of what he could have been causally acquainted with. This is where his laboratory notebooks and scientific instruments are particularly valuable to us. From them we are able to realize, e.g., that because his apparatus was not nearly sophisticated enough he certainly could not have confused things at the atomic level with things at the sub-atomic level. We are also able to establish — from interpreting observation reports alone! — exactly what "techniques of chemical analysis and synthesis" Dalton was familiar with. So if Dalton says, as he does, that "atoms" were irreducible given the known techniques, we can even repeat his experiments in order to better understand what particles he was dealing with.

In the last chapter I suggested, apropos of some remarks about how to discover the referent of a proper name, the following principle governing the use of a natural kind predicate within a linguistic community: for a natural kind predicate ' ϕ ' to be used by members of a linguistic community C to describe Ψ 's, it is necessary that there be some causal connection between Ψ 's and members of C, namely that between Ψ 's and what is believed about ϕ 's in C. As it stands this is a somewhat rough-and-ready principle, but the idea behind it seems to hold good. For Dalton to have propounded a theory, and hence to have had beliefs, about what we call atoms and molecules, and what he called "atoms", he must have been in a position to causally obtain knowledge of them through experiment. To take another case: for us to translate Mendel's 'bildungsfähig Element' as 'gene', it must be the case that genes are responsible for the phenomena which he observed in his experiments

on varieties of Pisum sativum and which he tried to interpret in his 1865 paper on plant hybridization.⁴⁷

This idea receives considerable support from the writings of Putnam and Kripke. In "Explanation and Reference," Putnam begins by arguing that what is important in discovering the extension of a physical magnitude term like 'electric charge' is knowing that there should have been an "introducing event" in which the term was correctly applied.⁴⁸ Such an event is one where there is a causal connection between the introducer of the term and what we recognize as an instance. By way of example, he asks the reader to imagine him standing next to Franklin as he performed his famous experiment with the kite. From this, he says, he is able to acquire the term 'electricity', and consequently the term 'electric charge'. He then extends his argument to natural kind predicates:

For natural kind words too, then, linguistic competence is a matter of knowledge plus causal connection to introducing events (and ultimately to members of the natural kind itself). And this is so for the same reason as in the case of physical magnitude terms; namely, that the use of a natural kind word involves in many cases membership in a 'collective' which has contact with the natural kind, which knows of tests for membership in the natural kind, etc., only as a collective.⁴⁹

The causal account of how the extension of a predicate is established suggests an important constraint on discovering what earlier scientists were talking about, and hence on interpreting their theories. In the case of natural kind predicates it is clear how the causal network is secured, for as I said before, such predicates apply to things in virtue of their physical properties. Despite Putnam's claims to the contrary, however, the causal account is not so obvious

in the case of non-natural kind predicates like physical magnitude terms. Putnam's discussion and example seem to slide from the physical magnitude term 'electric charge' to the natural kind predicate 'flow of electric charge', and thence to 'electricity'. In the next chapter I shall consider physical magnitude terms at greater length.

When Putnam published "Explanation and Reference" in 1973, he was an advocate of a "causal theory of reference", an expression the accuracy of which was challenged in the previous chapter. During the next three years he changed his mind and came to regard it as "a theory of how reference is specified,"⁵⁰ not of what reference is. With this new appellation we can well agree, for an important factor in discovering what the referent or extension of a term is, is how it came to be specified. As we have seen, this involves certain physical facts about the relation of the introducer or specifier to the world. But since there are such physical facts, one may wonder why it is that Quine excludes them in his purview of translation.

In the first section of this chapter we cited Quine as explaining the difference between indeterminacy of translation and underdetermination of physical theory by evidence using these words:

Consider, from this realistic point of view, the totality of truths of nature, known and unknown, observable and unobservable, past and future. The point about indeterminacy of translation is that it withstands even all this truth, the whole truth about nature. This is what I mean by saying that, where indeterminacy of translation applies, there is no real question of right choice; there is no fact of the matter even to within the acknowledged underdetermination of a theory of nature.⁵¹

The totality of physical facts, Quine is saying, fails to determine the correct translation. But there is a striking contrast between

such talk about "the whole truth of nature" and the restriction of this truth, in discussion of inscrutability and indeterminacy, to facts about behaviour. We agreed early on that theory is underdetermined by evidence and that the translation of observational sentences is determined by the behavioural facts. Suppose we now agree that the translation of observational sentences does not determine the translation of theoretical sentences. Even so, to get the required conclusion that the translation of theoretical sentences is not determined by the totality of physical facts, Quine needs the additional premise that the only physical facts relevant to translation are behavioural. But clearly the totality of behavioural facts forms only a small part of the totality of physical facts. So why is it that Quine thinks only the behavioural ones are relevant to translation?

We touched on what I think is the answer in Chapter 3. There we agreed to follow Quine in not requiring, in discussion of inscrutability, facts other than those which relate to the behaviour of language users. Quine begins the Preface to Word and Object as follows:

Language is a social art. In acquiring it we have to depend entirely on intersubjectively available cues as to what to say and when. Hence there is no justification for collating linguistic meanings, unless in terms of men's dispositions to respond overtly to socially observable stimulations. An effect of recognizing this limitation is that the enterprise of translation is found to be involved in a certain systematic indeterminacy.

Ten years later he reaffirms the same point:

Meanings are, first and foremost, meanings of language. Language is a social art which we all acquire on the evidence solely of other people's overt behaviour under publicly recognizable

circumstances. Meanings, therefore, those very models of mental entities, end up as grist for the behaviourist's mill. Dewey was explicit on the point: Meaning ... is not a psychic existence; it is primarily a property of behaviour." ⁵²

In one sense, of course, Quine is right to say that we acquire language from our observations of other people's overt behaviour. But it seems obvious that although we begin to acquire language in this way, we can progress by relating these observations to the wider body of physical information we are able to obtain about them and the world. In like manner we develop a theory of, say, quasars not just from observing their behaviour but by relating it to a wider body of physical theory. Similarly, when it comes to translation our parameters are fixed by our "theory of nature", by the totality of known physical facts, and these may outrun the behavioural ones.

This leads us back to the third stage of Quine's argument, to whether translation is indeterminate, as he says it is, or just underdetermined, as the rest of physical theory is. Føllesdal has claimed that the essential point for this stage, one "that nobody who has discussed Quine's views appears to have fully understood," is that "all the truths there are, are included in our theory of nature."⁵³ He goes on, "in our theory of nature we try to account for all our experiences. And the only entities we are justified in assuming are those that are appealed to in the simplest theory that accounts for all this evidence."⁵⁴ If what I have said is right, then some of our experiences — think here of Putnam standing next to Franklin — are only to be fully explained by considering the physical facts in addition to the behavioural ones. Our "theory of nature" is what it says it is, not a theory of behaviour! The conclusion, I suggest, is that translation, like the rest of physical theory, is underdetermined

by evidence, but has so far not been shown to be indeterminate.

In this section it has been maintained that there is no reason for thinking that Davidson's theory of radical interpretation will resolve all problems of determinacy, even given the diversity and complexity of the possible linguistic evidence. As for the constraint suggested by the causal account of how reference is specified, this only permits the scrutability of reference; it does not guarantee that we shall be able to capture the full sense of the term. Being able to discover what was causally responsible for the knowledge that Dalton, Mendel, Franklin, etc., had is not the same as being able to resolve all problems as to the meanings they attached to the terms they used. For we are relying on a theory of how reference is specified, and this is quite different from a theory of how meanings are to be attributed. Even under the theory of radical interpretation, then, the path from reference to meaning is narrow at best and might sometimes peter out. So it seems we are obliged to hold that scrutability of reference does not imply that translation is determinate. Our choice of a theory of truth for a language might to this extent be underdetermined by the physical evidence.

This recalls an argument from section (i). I said that if scrutability of reference did imply determinacy of translation, then indeterminacy of translation implies inscrutability of reference, in which case our being unable to determinately translate would preclude our assigning referents or extensions to alien terms. In this section we have seen that scrutability does not, in general, imply determinacy of translation. Consequently it cannot be said that failure to determinately translate implies inscrutability of reference. A theory

of truth for a language might be underdetermined, but this does not in itself give us a reason for thinking that we cannot discover what the names of the language refer to or what the predicates of the language have as their extensions. On the other hand, of course, we have not established the much stronger conclusion that we shall always be able to discover this. We have to take each case separately and weigh the evidence. In doing so we are with Quine when he says "knowledge, mind, and meaning are part of the same world that they have to do with, and ... they are to be studied in the same empirical spirit that animates natural science,"⁵⁵ though we might beg to differ over what those words mean.

CHAPTER 7 A BROADER PERSPECTIVE

Section (i): The Story So Far

In Chapter 2, I posed four interconnected questions which, I argued, a realist would have to answer in order to support his account of the growth of scientific knowledge. Having answered questions (3) and (4), and discussed how to interpret the language of previous theorists, we are now in a position to sum up our answer to the second, to the question

- (2) How can we discover which objects belong to the extension of a natural kind predicate as that predicate is used within a linguistic community?

The key to our answer is that we have to determine what beliefs members of the linguistic community who used the predicate had about things to which they thought it could be correctly applied. (I here confine myself to terms of past theories.) We need to look at what Johannsen (the biologist who coined the term 'gene') and those following him believed about what they called genes, at what our ancestors believed about whales, at what beliefs Bohr held about the things he called electrons in 1911, and so on. Of course, it will not always be the case that our predecessors used the same term that we do for a given natural kind (think here of Dalton calling molecules of gases "atoms"), although when we begin to interpret their theory or language we might at first assume that they do.

Before we actually begin our interpretation of the data we can state, as a general a priori constraint on possible attributions of extensions to predicates, that members of the linguistic community

under consideration must have been in an appropriate causal relation to things of whatever kind we decide they were applying a predicate to (a similar constraint holds for the attribution of referents to singular terms). For us to say that Bohr was talking about electrons, he must have performed experiments with them, or at least have been acquainted with the results obtained by others who had performed such experiments. Members of the natural kind — and here it is we who decide, on the basis of our own "theory of nature", what the members are — have to be (causally) "at the root of" beliefs if these beliefs are to be properly described as being about things of that kind. Here we see the importance to the realist of the connection between words and the world. In order for members of a linguistic community C to be interpreted as using a natural kind predicate: ' ϕ ' to describe ψ 's, it is necessary that there be some causal connection between ψ 's and members of C , namely that between ψ 's and what is believed about ϕ 's within C . This is a point rightly emphasised by those who argue for a causal account of reference.

Bearing this "causal constraint" in mind, we begin to interpret. The data on which we primarily concentrate are the sentences accepted as true by members of the linguistic community in question. If we wish to interpret a language, then our aim is to establish a theory of truth for that language; if we wish to interpret a scientific theory, then our aim is to establish a theory of truth for sentences of the scientific theory. With regard to the latter case, though, it must be remembered that any theory of truth for a part of a language must cohere with a theory of truth for the whole. As we have seen, the more theoretical sentences interrelate with the less theoretical to form the network of language.

To begin with, we have to assume that members of the community are correct, from our point of view, as often as possible; hence the assumption about their using a natural kind predicate to describe the same sort of thing that we use it to describe. In maximizing agreement in this way we attribute to them many of our own beliefs. Unless we thereby obtain some purchase on the notion of meaning we cannot begin interpretation proper.

First-order logic is attributed to members of the community. Working with those sentences always held true or always held false, and valid patterns of inference, we fit our orthodox logic to the language being interpreted. Having settled matters of logical form, we turn to the most observational sentences, i.e., to those whose truth-value bears the most obvious relation to changes in the environment. To give a theory of truth for a language is to state a procedure that will generate, for any sentence of the language, a T-sentence that is true. In the first place we gather evidence for those T-sentences where the object language sentence is most observational. We then proceed to the less observational. Here we are helped by the inferential links between more and less observational sentences. There is also likely to be a vast amount of primary and secondary linguistic information available, both in the case of interpreting a language as a whole and also in the case of interpreting a scientific theory. Finally, there is the other non-linguistic physical evidence of the kind I suggested would be of use, if taken together with the causal constraint, in interpreting Dalton's atomic theory of gases.

In the light of this further information and the suggested interpretations, it might prove necessary to revise, in certain

places, our original charitable assumptions. This in turn might lead us to alter some of our interpretations, and so on. In general there is no ground for assuming that there will be just one admissible theory of truth for a language on the basis of the available data (or even on the basis of all the possible data in Quine's sense). To return to the Quinean idiom, sentence translation may well turn out not to be determinate. It may even turn out to be indeterminate; though none of Quine's arguments show that it will. In themselves, neither is a cause for concern to the realist, for he is satisfied with optimal translation, translation that is consistent with all the evidence; he need not insist on there being perfect or unique translation. What would be a cause for his concern would be if such lack of determinacy implied that we cannot 'scrute' the reference or extension of terms of the object language. As I argued in the previous chapter, however, indeterminacy of translation of sentences and inscrutability of reference of terms are logically independent doctrines.

Applying this strategy to particular cases, the realist can hope to discover the extension of a natural kind predicate as it is used within a given linguistic community. He can hope to discover, for example, that Mendel's 'bildungsfähig Element' was understood by him to apply to those things which were responsible for the transmission of differentiating characteristics from one generation to the next. This might be expressed more formally as follows,

- (i) For any a, 'bildungsfähig Element' is true of a in Mendel's theory of inheritance if and only if a is responsible for the transmission of differentiating characteristics from one generation to the next.

Stating what the extension of a predicate is, is done through a metalanguage. Here we are reminded of the point made towards the end of Chapter 2, that in order to compare the extensions of relevant natural kind predicates from different scientific theories we must state what they are in the same meta- or background language.

Suppose that the realist wishes to compare Mendel's theory with Muller's. He will then be hoping to discover that the following is true,

- (ii) For any a, 'gene' is true of a in Muller's theory of genetics if and only if a is responsible for the transmission of differentiating characteristics from one generation to the next.

Assuming that (i) and (ii) are expressed in the same language, it follows that what Mendel called 'bildungsfähigen Elemente', Muller called 'genes'; or, more formally, that

- (iii) For any a, 'bildungsfähig Element' is true of a in Mendel's theory of inheritance if and only if 'gene' is true of a in Muller's theory of genetics.

And this will be so even though many of the beliefs held by Mendel and Muller about their subject-matter might be different. That is, it will be so even though the senses they attached to the terms they used might be different.

The complex path leading to (iii) constitutes an answer to a particular instance of question (1) in its material mode. This is the primary question in the realist's account of the growth of science. If (i) and (ii) are true, then Mendel and Muller were talking about the same things. Perhaps the most difficult task is establishing that what Mendel and Muller regarded as the differen-

tiating characteristics of a species were more or less the same. Naturally this would itself involve interpretation that would suggest statements of the same form as (i) and (ii). But notice that 'differentiating characteristics' is a less theoretical, more tractable term than 'gene' (or 'bildungsfähig Element'). We might therefore expect to be able to make more use of laboratory notebooks, observation reports, and also of the kind of information afforded by examining the apparatus used by Muller when considered along with the causal constraint.

In general, how successful the realist will be if he pursues the strategy I have been arguing for as an explanation of the growth of science is not something that can be specified in advance; it is a contingent matter. There may be many cases where he cannot establish the extensions of predicates from previous scientific theories. And even where he can it might turn out that subsequent or competing theories are not about the same things.

So far in this thesis I have concentrated on cases where the relevant scientific predicate is a natural kind predicate and where one's intuitive view is that subsequent theories are about the same things. I have argued for a number of philosophical points which make sense of this intuition. By way of conclusion I wish to make some remarks on two cases of rather different sorts, one where we now say that there is nothing of the kind postulated by earlier theorists, and one where the relevant predicate is not a natural kind predicate. Though my discussion of these cases will not be extensive, I hope that the way they contrast with those considered so far will make clear the limits of the foregoing arguments and thereby suggest in what direction further work might proceed.

Section (ii): 'Phlogiston'

The first case turns on the difference between two sorts of error we might wish to ascribe to our ancestors: a predicate they used failing to have any extension at all according to our conceptual scheme, and its having an extension but their holding false beliefs about what they correctly applied it to. I have already mentioned several cases which appear to be of the latter kind — Bohr thought that at all times electrons have precise positions and momenta, Muller believed that genes were composed of proteins, Ptolemy held that the other planets revolved around the earth, and so on. Less attention has been paid to those cases where we now assign a null extension to predicates from previous scientific theories — from our point of view there is no phlogiston, no luminiferous ether surrounding the earth, no caloric surrounding atoms, and so on. Is there a clear dividing line between these two sorts of cases? How do we decide between them?

We have to be careful here to distinguish between questions of the form 'What, according to us, is the extension of the predicate ' ϕ '?' and questions of the form 'Did members of linguistic community C succeed in describing anything when they used the term ' ϕ '?'. According to us, the predicate 'phlogiston' has null extension — there is no such thing as phlogiston, those who believed that there was were mistaken. Nevertheless, chemists during most of the 18th century described the results of some of their experiments using the predicate, and in so doing were thought by their contemporaries to have made true statements. Thus, suppose that one such chemist offers the following report of a familiar laboratory observation:

(OR1) On addition of oil of vitriol to granulated zinc, evolution of phlogiston was observed.

Such a report would have been accepted as true by other chemists of the time. To us it seems obvious what the chemist in fact observed, namely, the evolution of hydrogen on addition of sulphuric acid to granulated zinc. (We can interpret their laboratory notebooks.) Yet if we fail to distinguish between the above questions — if we fall victim to "the chauvinism of time"¹ — then since there is (tenseless) no such thing as phlogiston, there can be no such event as an evolution of phlogiston, and so the report cannot be said to describe any event at all.

Such failure to distinguish leads to two problems. The first is that not being able to assign the value 'true' to many of their observation reports conflicts with the principle of charity. We might now say that there is no such thing as phlogiston, but if we hold that it follows from this that the predicate lacked an extension for early 18th century chemists, then our interpretation of their science will diverge in countless places from their own understanding. The sort of generalization which reports like (OR1) will tend to support is

(x)(if x is a member of the community of early 18th century chemists then (x holds 'OR1' true if and only if x observes an evolution of hydrogen when sulphuric acid is added to granulated zinc)).

This in turn supports the T-sentence

'OR1' is true in the language of early 18th century chemistry for a speaker x if and only if x observes an evolution of hydrogen when sulphuric acid is added to granulated zinc.

But neither the generalization nor the T-sentence are supported if we refuse to assign an extension to the predicate 'phlogiston' as used by them.

The second problem is related to the first and can also be stated briefly. If such observation reports fail to describe anything at all, then the widespread agreement by chemists of the time over what truth-value they are to be assigned must strike us as nothing short of miraculous. How could they agree if there was not anything there for them to correctly apply their predicates to? This point is made by Nick Jardine in a recent paper,² where he levels it against a principle for making retrospective assignments of extension which Putnam calls the "principle of benefit of the doubt".³ According to Putnam's principle, we are permitted to identify the extension of a predicate ' ψ ' of our science with that of a simple predicate ' ϕ ' (like 'gene' or 'atom') of a past science provided only that the descriptions which protagonists of the past science used to characterize ϕ 's would, when "reasonably reformulated", characterize ψ 's. This principle, says Putnam, would allow us to "assign a referent to 'gravitational field' in Newtonian theory from the standpoint of relativity theory (though not to 'ether' or 'phlogiston'); a referent to Mendel's 'gene' (sic) from the standpoint of present-day molecular biology; and a referent to Dalton's 'atom' from the standpoint of quantum mechanics."⁴ What Jardine claims is that accepting this principle requires us to acknowledge, in cases like 'phlogiston', that there can be miracles! He also sees another difficulty in its acceptance. The principle forces us into a dilemma when it comes to interpreting simple predicates of a past science. Either we must equate its extension

with that of a simple predicate of our own science, or we must admit that it has null extension. It forces us to accept that there is a clear dividing line between the two sorts of cases.

Seeking to avoid the postulation of miracles, Jardine is led to argue "that it is a precondition for assignment of extension to the predicates of a past science that that assignment be such as to make many such observation reports [as (OR1)] come out true."⁵ We are well placed to agree with Jardine here, for not only do we too wish to avoid talking of miracles, we also have to defend a principle of charity underlying our theory of interpretation. Let me now explain how what I have said earlier in this thesis meets Jardine's precondition.

The theory of reference for natural kind predicates that I adumbrated in Chapter 5 was a theory of what it is for an object a to belong to a natural kind predicate 'Q' as that predicate is used within a linguistic community C. In keeping with this, specifying the extensions of predicates in statements (i) and (ii) from the last section was accomplished using a three-place relation holding between a term, a set of objects and a theory. If it should turn out that, say, certain occurrences of hydrogen satisfy a suitable majority of those descriptions which could be consistently attributed to phlogiston on the basis of what was believed by early 18th century chemists about phlogiston, then this theory of reference will reckon those instances to form part of the extension of 'phlogiston' as used by early 18th century chemists. The realist could thus hope to discover that the following is true:

- (iv) For any a, 'phlogiston' is true of a in early 18th century chemistry if a is an occurrence of hydrogen resulting from

the addition of sulphuric acid to granulated zinc.

Another observation report on which Georgian chemists also agreed is that phlogiston is evolved when certain substances are burnt in air. What in fact happens, according to us, is that oxygen combines with the substances. Using the same procedure as was used to discover the truth of (iv), the realist could hope to establish as true

- (v) For any a, 'phlogiston' is true of a in early 18th century chemistry if a is an occurrence of oxygen which combines with a substance when it is burnt in air.

Combining (iv) and (v) he then obtains

- (vi) For any a, 'phlogiston' is true of a in early 18th century chemistry if a is an occurrence of matter which is either hydrogen resulting from the addition of sulphuric acid to granulated zinc, or oxygen combining with a substance when it is burnt.

Now as a matter of fact Georgian chemists believed that several other experiments led to the production of phlogiston. Once these have been accounted for the realist, as Jardine also notes,⁶ will specify the extension of their simple predicate using a complex disjunctive predicate. That is, what he finally comes to accept as true will be something of the form

- (vii) For any a, 'phlogiston' is true of a in early 18th century chemistry if and only if a is an occurrence of matter which is either hydrogen resulting from the addition of sulphuric acid to granulated zinc, or oxygen combining with a substance when it is burnt, or ...

Such a retrospective assignment of extension ensures that (OR1), and

other well established reports like it, are assigned the value 'true' in the language of 18th century chemists.

Accepting (vii) also enables the realist to avoid the dilemma forced on one who accepts Putnam's "principle of benefit of the doubt". He is free to recognize that contemporary chemists have no one simple predicate to which they assign the same extension as was assigned by early 18th century chemists to their simple predicate 'phlogiston', without thereby committing himself to the conclusion that their predicate has a null extension. More generally, it is to be expected that the theory of interpretation which I have presented, based as it is on a cluster theory of reference, will always urge us to assign a non-empty extension to predicates of previous theories provided we can detect a consistent use of the predicate. If we repeatedly end up with inconsistent assignments of beliefs to users of the predicate, then we have no reason for thinking that there are things to which they believed it could be correctly applied. Jardine suggests that Gassendi's predicate 'atom' might turn out to be such a case;⁷ perhaps 'phlogiston' as used by late 18th century chemists would be another.⁸ But where a predicate is used consistently within a community there are good intuitive and methodological reasons for assigning an extension to it relative to that community.

In between these two possibilities there appears to be a third: we may be unsure, given the vagaries of interpretation, whether or not a predicate was used consistently. We might be able to vary assignments of truth-values to sentences and so specify different extensions for their predicates. Perhaps the earlier theorists were not as truthful as we at first charitably assumed; perhaps the

evidence at hand will be just too thin to resolve the matter. If this is a possibility then there is no clear dividing line between cases involving false beliefs and those involving non-instantiated predicates.

The account I have given of how we should assign extensions to predicates of past theories leads to the following picture of the growth of scientific knowledge. Suppose there is a theory T^1 which contains a natural kind predicate 'P' and which comes to be replaced by theory T^2 containing natural kind predicate 'Q'. The question arises as to whether those things described as P's by users of T^1 are the same as those described as Q's by users of T^2 . According to my account, an affirmative answer should be given to this question when a suitable majority of those descriptions believed to be true of P's by users of T^1 also constitutes a suitable majority of those descriptions believed to be true of Q's by users of T^2 . As an example of this I suggested, in the previous section, that what Muller described as genes, Mendel had described as bildungsfähigen Elemente; the 'suitable majority of descriptions' being 'whatever is responsible for the transmission of differentiating characteristics from one generation to the next'. In this section I have considered the case of phlogiston theory. After Lavoisier and the discovery of oxygen, phlogiston theory was given up. According to post-18th century chemists, no one thing satisfied a suitable majority of those descriptions which they could consistently attribute to phlogiston on the basis of what they believed. Probably the main reason for their having adopted this view is that as experimental techniques improved towards the end of the 18th century it was demonstrated that, since all metals gained weight on calcination,

phlogiston, which was thought to cause calcination by being evolved, must have negative weight. But the idea of an element having negative weight seemed then, as it has ever since, to be of dubious intelligibility. There was no question, moreover, of this difficulty being overcome. It was not just a matter of one false belief amongst many true ones. Phlogiston's having negative weight was a consequence of the theory itself and not something which could be shelved or left for future science to resolve. The theory was abandoned and no one simple predicate with the same extension was subsequently used.

None of this, however, prevents us from interpreting the language, and making sense of the actions, of early 18th century chemists in the way suggested. The realist need not be a chauvinist. He can make sense of errors on the part of earlier scientists. He can also render intelligible such statements as 'In the dark ages they talked about witches, although in fact such people were mainly schizophrenics', but to discuss this would take me too far afield.

Section (iii): 'Mass'

The final example I want to discuss leads us, in a sense, back to our starting point, for it is one of the main examples considered by Feyerabend in arguing against accounts of the growth of science which were based on the notion of the reduction of an early theory to a later one. From Chapter 1 we recall that he discusses a number of purported reductions of this kind which Nagel claimed to be in accordance with the conditions of deducibility — 'the laws of the reduced theory must be logically derivable from the reducing theory together with certain coordinating definitions' — and meaning

invariance — 'the meanings of the theoretical terms contained in the derived laws are the same as those of the terms as they occur in the reduced theory'. As an example of reduction where meanings are supposed to have remained invariant, Feyerabend considers the replacement of classical mechanics by relativity theory, concentrating on the term 'mass' as it occurs in each.⁹ The essential difference between the two theories is that classical mechanics assumes that the mass of a particle is constant whereas in relativity theory it is said to be proportional to a frame of reference. According to Feyerabend there is a change of meaning here because classical mass is a property of an object itself whereas relativistic mass is a relation, and because (apparently) incompatible equations about mass hold in the two theories.

I noted at the beginning of Chapter 2 certain difficulties which arise with Feyerabend's position, particularly with regard to claims about the incompatibility of incommensurable theories. In part these were traced back to a general scepticism about the notion of meaning. Aiming to avoid this, while at the same time preserving a sense of scientific progress, I turned to what a realist might say in response. The nub of the realist's view is that successive scientific theories are often theories of the same kinds of thing. Scientists usually act in accordance with this view when formulating new theories and, says the realist, science progresses as a result. We have seen that the realist cannot fully abjure meaning — he makes use of the notion in explaining how we come to understand what theories are about. But he is not committed to the condition of meaning invariance. All that a realist need claim remains invariant are the extensions of predicates.

In discussing the realist's general view I concentrated on natural kind predicates. One interesting feature of them is that they do not usually admit of simple, precise definitions. To connote this feature I adopted the expression 'cluster term', which Putnam uses to describe these predicates. The point is that natural kind predicates are defined using a cluster of properties, not all of which need be true of an object in order for it to be reckoned one of the kind. Of some relevance to the present context is the fact that Putnam's intention, in first introducing the expression, is to use it to characterize terms like 'energy', 'mass', 'temperature', and so on which frequently occur in scientific laws. These he calls "law-cluster concepts".¹⁰ According to Putnam,

Law-cluster concepts are constituted not by a bundle of properties as are the typical general names like 'man' and 'crow', but by a cluster of laws which, as it were, determine the identity of the concept. The concept 'energy' is an excellent example of a law-cluster concept. It enters into a great many laws. It plays a great many roles, and these laws and inference roles constitute its meaning collectively, not individually.¹¹

This feature, which goes to make such terms law-cluster concepts, leads Putnam to draw the following conclusion about theory change,

In general, any one law can be abandoned without destroying the identity of the law-cluster concept involved, just as a man can be irrational from birth, or can have a growth of feathers all over his body, without ceasing to be a man. Applying this to our example — 'kinetic energy' = 'kinetic' + 'energy' — the kinetic energy of a particle is literally the energy due to its motion. The extension of the term 'kinetic energy' has not changed. If it had, the extension of the term 'energy' would have to have changed. But the extension of the term 'energy' has not changed. The forms of energy and their behaviour are the same as they always were, and they are what physicists talked

about before and after Einstein. On the other hand, I want to suggest that the term 'energy' is not one of which it is happy to ask, What is its intension? The term 'intension' suggests the idea of a single defining character or a single defining law, and this is not the model on which concepts like energy are to be construed.¹²

If Putnam is right here then this is of great importance to the realist in the face of the present problem. It suggests that he can accept that the meaning of the term 'mass' might not have stayed the same in the change from classical mechanics to relativistic mechanics but that the extension of the term has. What might be a cause for scepticism, though, is Putnam's bold claim that "the forms of energy and their behaviour are the same as they always were." What exactly does he mean by "the forms of energy and their behaviour"? And how are we to decide what the behaviour of something is if not through consideration of the laws it is said to obey? Before discussing the problem further, let me make some more general remarks.

One important difference between natural kind predicates and law-cluster terms is that the former are used to refer to things of a kind whereas the latter are used to express properties that things have. That is, law-cluster terms are typically used to express measurements. For this reason they are more correctly said to be two(or more)-place predicates; natural kind predicates, on the other hand, are one-place. The classical mass of a particle, e.g., is expressed by a functional relation between that particle and a number. This contrast underlies the one Putnam expresses in terms of a difference between "a bundle of properties" and "a cluster of laws".

Since terms for measurements are two(or more)-place predicates, their extensions consist of ordered pairs(or n-tuples). It would appear, then, that the realist's claim of "reference invariance", applied to them, amounts to the claim that the law-cluster predicates of successive theories have the same sets of ordered pairs(n-tuples) in their extensions. Where before we asked questions like whether the extension of Newton's 'planet' was the same as the extension of ours, we now have to ask whether the values assigned to, say, 'the mass of planet x' within Newton's theory of gravitation correspond to values assigned by any predicate of ours of the form 'the Ψ of planet x'. It might be the case that we do not believe some things about ' Ψ ' which Newtonians believed about 'mass', but in point of extension they have to be the same.

This contrast between predicates the extensions of which are things and predicates the extensions of which are ordered n-tuples suggests a difference in the intuitive appeal of the realist's account of theory change for the different types of case. In the first the realist can say, e.g., that Mendel and Muller were both talking about just these things — genes, or that Ptolemy and Newton were talking about just this planet — Venus. In the second, though, he seems to be limited to saying things like that Newton and Einstein would both have assigned the value 6 to this object. The measurements have to be made; the kinds of thing that are measured are already there.

The difference here can be made clearer if we concentrate on the notion of a causal relation. This formed an important part of my account of how we can discover what the extension of a natural kind predicate is. I argued that in order for members of some

linguistic community to be said to use a natural kind predicate ' ϕ ' to describe ψ 's, they (or their ancestors) must have been causally related to ψ 's. Numbers, however, do not stand in causal relations to people (or to anything for that matter). Ptolemy and Newton might both have observed the planet Venus — it reflected light from the sun onto their retinas — just as Mendel and Muller both observed the effects of genes. But there is no sense in saying that Newton or Einstein could have observed 6 or its effects (though of course they could have observed a pointer indicating '6' on a scale). This difference, it seems to me, largely accounts for the special problems associated with theory change where the relevant predicate is a law-cluster term.¹³ The connection between words and the world is more complex.

Returning to the 'mass' example, this has been discussed at length by Hartry Field in a paper entitled "Theory Change and the Indeterminacy of Reference."¹⁴ As the title suggests, Field thinks it has something to do with a doctrine of Quine's (exactly which doctrine is a point discussed below). This is made clear from the outset where he states his general thesis as being "that considerations about scientific revolutions show that many scientific terms are referentially indeterminate — there is no fact of the matter as to what they denote (if they are singular terms) or as to what their extension is (if they are general terms)."¹⁵ His sole argument in support of this thesis rests on the example we have been discussing.

The essence of Field's claim is this. In relativity physics there are two properties — proper mass and relativistic mass — each of which resembles classical mass, though in different respects.

Thus, the relativistic mass of a particle, when multiplied by its velocity, is equivalent to its momentum; this is not so for proper mass. The proper mass of a particle, on the other hand, is the same in all frames of reference; its relativistic mass is not. So no matter which property is chosen as the "successor" to classical mass, classical physics must be judged wrong in some respect or other when viewed from the standpoint of relativity physics. Moreover, according to Field there is no fact of the matter as regards which choice should be made. His conclusion is that we have an instance of "the indeterminacy of reference".

It seems to me that the initial reaction to this argument by anyone with some understanding of modern physics would be that Field is plainly wrong. Almost every modern textbook on physics goes to some length to point out that what Newton called 'mass' modern physicists call 'proper mass' or 'rest mass'. But the question remains, what substance can be given to this claim? In a recent reply to Field, John Earman holds that there is an "exact parallelism" between some central principles of Newtonian and special relativistic mechanics.¹⁶ To see the parallelism it is necessary to formulate both theories in what Earman calls "a four-dimensional, intrinsic (i.e., coordinate free) form."¹⁷ The three principles central to Newtonian mechanics can then be written as

$$(N1) \ m_n \text{ is a scalar invariant}$$

$$(N2) \ P_n = m_n V_n$$

$$(N3) \ F_n = m_n A_n$$

where m_n , P_n , V_n , F_n , A_n are, respectively, the Newtonian mass, the Newtonian four-momentum, four-velocity, four-force and four-acceleration. Earman's point is that in the special theory of relativity

"there are exact analogues (R1), (R2), (R3) of (N1), (N2), (N3) with proper mass m_0 in place of m_n , the relativistic four-momentum P_r in place of P_n , etc."¹⁸ So it would seem that while Field was right to say that in the usual formulation of the special theory, proper mass multiplied by velocity does not equal momentum, the theory can be written in an alternative "four-dimensional, intrinsic form" in which the equation does hold.

Let us accept that Earman is right about there being this "exact parallelism". Why should it convince us that classical mass is to be identified with proper mass? Earman's reply relies on its being the case that (N1) - (N3) and (R1) - (R3) occupy positions of great importance within their respective theories. He claims that "if any tenets are central to the Newtonian concept of mass and the relativistic concept of proper mass, it is (N1) - (N3) and (R1) - (R3) respectively."¹⁹ What is philosophically interesting about this is the idea that in cases of theory comparison where the relevant predicate is a law-cluster predicate, we should pay particular attention to the structure of the laws stated in the theories. This suggests a further criterion which could be used in resolving problems arising in such cases. We might phrase it as follows: where there is a prima facie choice over which predicate from one theory is to be said to have the same extension as a predicate from another, choose the one which would best preserve the mathematical structure between the theories. It will still be the case that, as Field says, classical physics will be judged wrong from the standpoint of relativity physics, but if Earman is right then we can at least see in the latter a successor to the former.

What then becomes of indeterminacy? Even with the added criterion, we have no guarantee that we shall always be able to determinately translate statements of Newton's, or, more generally, that we shall always be able to pick out the extension of a law-cluster predicate from another theory. But neither from the fact that translation is not determinate, nor even from the fact that it can be indeterminate, do we get that reference is inscrutable. This is where Field goes wrong. We might not be sure how to translate a Newtonian assertion like 'The mass of a particle is equal to twice its kinetic energy divided by the square of its velocity', even after we have appreciated what Earman has to say. We might even concede that there is no fact of the matter as regards its translation. Yet none of this would show the extension of Newton's 'mass' to be inscrutable. So in one sense Field is right — theory change might result in there being no fact of the matter as to how to translate — but this is not to say that there is "indeterminacy of reference".

Lastly, what about Feyerabend's views? He too has an objection to identifying the extension of 'classical mass' with that of 'proper mass'. In Chapter 1, I quoted him as saying "although both may have the same numerical value, they cannot be represented by the same concept."²⁰ If this is intended to express the point that not everything said of 'mass' in classical physics is accepted as true of 'proper mass' in relativity physics, then this might well be accepted by a realist. But this very claim is one that presupposes that the two theories are commensurable, a presupposition of which the realist alone can make sense.

Notes for Chapter 1

N.B. When emphasis occurs in a quotation, it is to be assumed, throughout the thesis, that it appears in the original, unless there is some note to the contrary.

1. J. Locke, Essay Concerning Human Understanding, Bk. 4, ch. 3, 25.
2. In particular, see the Preface, Bx-Bxiv.
3. I. Kant, Critique of Pure Reason, Introduction, section V.
4. For further detail see M. Hesse, Models and Analogies in Science.
5. M. Hesse, The Structure of Scientific Inference, p.286.
6. See, e.g., his early On the Definition of Mass.
7. E. Mach, "On the Economical Nature of Physical Inquiry," in his Popular Scientific Lectures.
8. R. Harre, The Philosophies of Science, p.75.
9. H. Poincaré, Science and Hypothesis.
10. See the remarks in his "Intellectual Autobiography," The Philosophy of Rudolf Carnap, ed. P. Schilpp, pp.15, 77-8.
11. See C. Hempel and P. Oppenheim, "Studies in the Logic of Explanation," Philosophy of Science, 15, 1948, particularly section 7; C. Hempel, "The Theoretician's Dilemma: A Study in the Logic of Theory Construction," reprinted in his Aspects of Scientific Explanation; and F. Suppe, "The Search for Philosophic Understanding of Scientific Theories," The Structure of Scientific Theories, ed. F. Suppe, section II.
12. For a possible axiomatization of T see J. McKinsey, A. Sugar, and P. Suppes, "Axiomatic Foundations of Classical Particle Mechanics," Journal of Rational Mechanics and Analysis, 2, 1953.
13. This is not intended as a complete definition. Further specification of the gas in which the substance burns would, for example, normally constitute part of the definition. Clearly though this only affects the content, not the form, of the definition.
14. R. Carnap, "Testability and Meaning," section 5, reprinted in Readings in the Philosophy of Science, ed. H. Feigl and M. Brodbeck.
15. Cf. *ibid.*, section 10.
16. K. Popper, "Three Views Concerning Human Knowledge," reprinted in

Notes for Chapter 1 (cont.)

- his Conjectures and Refutations, pp.118-19; N. Goodman, Fact, Fiction and Forecast, pp.40-1.
17. See, e.g., D.H. Mellor, "In Defence of Dispositions," Philosophical Review, vol. 83, 1974.
 18. P. Bridgman, The Logic of Modern Physics.
 19. Ibid., pp.6, 23-4.
 20. C. Hempel, "A Logical Appraisal of Operationalism," reprinted in his Aspects of Scientific Explanation.
 21. "Testability and Meaning," op. cit., p.47.
 22. C. Hempel, Fundamentals of Concept Formation in Empirical Science.
 23. N. Campbell, Physics: the Elements, reprinted as Foundations of Science: The Philosophy of Theory and Experiment, ch. VI.
 24. "The Search for Philosophic Understanding," op. cit., pp.25-6. A similar view is expressed by E. Nagel, The Structure of Science, p.106. For further discussion of correspondence rules, see Nagel, *ibid.*, pp. 97-105.
 25. "Testability and Meaning," op. cit., p.63. In this paragraph I draw on section 8 of Carnap's paper, and on his Philosophical Foundations of Physics, pp.225-6.
 26. The Structure of Science, p.118.
 27. In contrast to both instrumentalism and realism is what Nagel calls "descriptivism"; "according to it, a theory is a compendious but elliptic formulation of relations of dependence between observable events and properties," *ibid.*, p.118. This position falls with the abandonment of the view that correspondence rules are explicit definitions. It is also worth noting that in "The Theoretician's Dilemma," op. cit., Hempel distinguishes two versions of what I have called instrumentalism depending on what ontological commitments are being opposed.
 28. The Structure of Science, p.133.
 29. Ibid., p.143.
 30. Ibid.
 31. Ibid.
 32. "The Theoretician's Dilemma," op. cit., p.222.
 33. The Structure of Science, p.342.
 34. J. Kemeny and P. Oppenheim, "On Reduction," Philosophical Studies, vol. 7, pp.6-19.

Notes for Chapter 1 (cont.)

35. The Structure of Science, p.345; emphasis added.
36. Ibid., pp.353-54; emphasis added.
37. Probably the clearest statement of Popper's views on these issues is his "The Demarcation Between Science and Metaphysics," in The Philosophy of Rudolf Carnap, op. cit. Unlike the early positivists, Popper did not wish to maintain that traditional metaphysical questions are meaningless, but only that they are non-scientific.
38. K. Popper, The Logic of Scientific Discovery, Preface to the First English Edition.
39. "Three Views Concerning Human Knowledge," op. cit., p.115; italics deleted from original.
40. The Logic of Scientific Discovery, section 85, fn.*2. See also I. Lakatos, "Falsification and the Methodology of Scientific Research Programmes," Criticism and the Growth of Knowledge, ed. I. Lakatos and A. Musgrave, p.119, esp. fn.6.
41. The Logic of Scientific Discovery, p.278.
42. Ibid., section 12, fn.*1. For more general criticism, see "Three Views Concerning Human Knowledge," op. cit., section 5.
43. "Three Views Concerning Human Knowledge," op. cit., p.117.
44. Ibid., p.118; my emphasis.
45. Ibid., p.119.
46. P. Feyerabend, "Problems of Empiricism," Beyond the Edge of Certainty, ed. R. Colodny, p.214.
47. P. Achinstein, Concepts of Science, p. 160.
48. N. Hanson, Patterns of Discovery, p.19, etc.
49. N. Hanson, "Observation and Interpretation," Philosophy of Science Today, ed. S. Morgenbesser.
50. H. Putnam, "What Theories are not," reprinted in his Mathematics, Matter and Method, Philosophical Papers: Volume 1.
51. Ibid., p.218.
52. This point is made by F. Suppe, "On Partial Interpretation," Journal of Philosophy, 68, 1971, pp.57-76.
53. C. Hempel, "The Meaning of Theoretical Terms: A Critique of the Standard Empiricist Construal," Logic, Methodology and the Philosophy of Science, vol. 4, p.370.

Notes for Chapter 1 (cont.)

54. W. Quine, "Two Dogmas of Empiricism," reprinted in his From a Logical Point of View.
55. Ibid., p.41.
56. P. Feyerabend, "Explanation, Reduction and Empiricism," Minnesota Studies in the Philosophy of Science, vol. 3, ed. Feigl, Maxwell.
57. Ibid., p.33.
58. Ibid., p.28.
59. For further detail, see ibid., pp.46ff. Note that (a) Feyerabend is not drawing attention to the inconsistency of Newton's theory and Galileo's law, but to the inconsistency of some consequences of the two in a common domain of validity; (b) as Feyerabend observes, Duhem discussed the example in detail in his The Aim and Structure of Physical Theory, chapters 9 and 10.
60. "Problems of Empiricism," op. cit., p.168. He also considers the purported reducibility of classical mechanics to quantum theory, "Problems of Microphysics," Frontiers of Science and Philosophy, ed. R. Colodny. As a further example, one mentioned in section (ii), it can be shown that whereas Kepler's third law states that the ratio of the squares of the periods of any two planets to the cubes of the semimajor axes of their respective orbits equals a constant, a corollary of Newton's theory is that the same ratio varies according to the masses of the two planets.
61. "Problems of Empiricism," op. cit., pp.168ff.
62. Ibid., p.169.
63. Ibid.
64. The Structure of Science, p.352.
65. As a number of commentators have pointed out, they use the term 'incommensurable' to cover much else besides. In fact I can find no precise statement by either of them which is equivalent to the "definition" I have given. I have chosen it nevertheless because it captures the spirit of much that they want to say and also represents the strongest challenge to orthodox positivist, empiricist and realist views.
66. "Explanation, Reduction and Empiricism," op. cit., p.29.
67. Ibid.
68. T. Kuhn, The Structure of Scientific Revolutions.

Notes for Chapter 1 (cont.)

69. Ibid., ch. 10.
70. W. Sellars, "The Language of Theories," in Current Issues in the Philosophy of Science, ed. H. Feigl and G. Maxwell.
71. G. Maxwell, "Scientific Methodology and the Causal Theory of Perception," Problems in the Philosophy of Science, ed. I. Lakatos and A. Musgrave.
72. D.H. Mellor, "Physics and Furniture," American Philosophical Quarterly, 1970.

Notes for Chapter 2

1. P. Feyerabend, "On the 'Meaning' of Scientific Terms," Journal of Philosophy, vol. 62, 1965. The strong alternative seems to have been what he held in such early papers as "Explanation, Reduction, and Empiricism," Minnesota Studies in the Philosophy of Science, vol. III, and "How to be a Good Empiricist — a Plea for Tolerance in Matters Epistemological," reprinted in The Philosophy of Science, ed. P. Niddich.
2. "On the 'Meaning' of Scientific Terms," op. cit.
3. Several commentators have drawn attention to this point. See, e.g., D. Shapere, "Meaning and Scientific Change," Mind and Cosmos: Explorations in the Philosophy of Science, ed. R. Colodny, and P. Achinstein, Concepts of Science.
4. "Problems of Empiricism," Beyond the Edge of Certainty, ed. R. Colodny, p.168.
5. "Explanation, Reduction, and Empiricism," op. cit., pp.28-9.
6. W. Quine, "Two Dogmas of Empiricism," reprinted in his From a Logical Point of View.
7. One attempt in this direction is F. Suppe, "On Partial Interpretation," Journal of Philosophy, 68, 1971. Nevertheless, he too seems to accept the judgment expressed in my next sentence, see his "The Search for Philosophic Understanding of Scientific Theories," The Structure of Scientific Theories, ed. F. Suppe, p.203.
8. Quine is, of course, the philosopher usually characterized as holding this view. Rather different arguments leading to much the same end are presented by J. Austin, "Are There A Priori Concepts?" reprinted in his Philosophical Papers.
9. Compare, e.g., some of the papers in Semantics of Natural Languages, ed. D. Davidson and G. Harman, and some of those in Mind and Language, ed. S. Guttenplan.
10. English translation "On Sense and Reference," Philosophical Writings of Gottlob Frege, trans. and ed. P. Geach and M. Black.
11. "Function and Concept," Philosophical Writings of Gottlob Frege, op. cit.; in particular see p.29 and fn.
12. "On Sense and Reference," op. cit., p.56.
13. Ibid., p.61.

Notes for Chapter 2 (cont.)

14. For further discussion see M. Martin, "Referential Variance and Scientific Objectivity," and "Ontological Variance and Scientific Objectivity," British Journal for the Philosophy of Science, vol. 22, 1971, and vol. 23, 1972, respectively.
15. H. Putnam, "Explanation and Reference," reprinted in his Mind, Language and Reality, Philosophical Papers: Volume 2, p.197.
16. Ibid., see also "The Meaning of 'Meaning'," in the same volume.
17. K. Parsons, "A Criterion for Meaning Change," Philosophical Studies, vol. 28, 1975.
18. J. Dalton, A New System of Chemical Philosophy, Part I, p.218.
19. For further discussion see the papers of Dalton's reprinted in Foundations of the Atomic Theory, Alembic Club Reprint, no. 2, Foundations of the Molecular Theory, Alembic Club Reprint, no. 4, and Parsons' "A Criterion for Meaning Change," op. cit.
20. Reprinted in Foundations of the Molecular Theory, op. cit.
21. See, e.g., A New System of Chemical Philosophy, Part I, p.163.
22. Foundations of the Molecular Theory, op. cit., p.28, fn.2.
23. A. Tarski, "The Concept of Truth in Formalized Languages," reprinted in his Logic, Semantics, Metamathematics.
24. "Explanation and Reference," op. cit., p.198.
25. Ibid. I have replaced Putnam's ' T^1 ' and ' T^2 ' by ' T ' and ' T' '.
26. "A Criterion for Meaning Change," op. cit., p.372. N.B. I have added two sets of quotation marks missing from the criterion as given but explicitly assumed by Parsons in her discussion.
27. Ibid.
28. Ibid., p.391. I have replaced Parsons' 'for any a ' by ' (x) ', occurrences of a by x , and her 'if and only if' by ' \equiv '. Also, her discussion is purely schematic; she does not suggest a particular interpretation for ' P ', ' Q ', ' T^1 ', and ' T^2 '.
29. Ibid., p.378.
30. Ibid.
31. Ibid.
32. Ibid., p.379.

Notes for Chapter 3

1. W. Quine, "Ontological Relativity," in W. Quine, Ontological Relativity and Other Essays, pp.48-9; hereafter this article is referred to as "O.R."
2. "O.R.", p.50.
3. W. Quine, "On the Reasons for Indeterminacy of Translation," Journal of Philosophy, vol. 67, 1970, pp.178-83.
4. Ibid., p.179.
5. For exposition I draw on Ontological Relativity and Other Essays, pp.1-6,30-35; Word and Object (hereafter WO), pp.51-54,68-79. Quine defines occasion sentences as those "which command assent or dissent only if queried after an appropriate prompting stimulation," WO, pp.35-36.
6. "O.R.", p.30. In WO, section 12, "the rabbit fusion" and "manifestation of rabbithood" are added as further alternatives; these will be considered towards the end of section (iii).
7. WO, section 15.
8. "O.R.", p.26.
9. Ibid.
10. Quine's assumption that the behavioural facts exhaust the physical facts will be discussed in Chapter 6, section (iii).
11. "O.R.", p.27.
12. "O.R.", p.47.
13. Ibid.
14. W. Quine, "Replies," Words and Objections, ed. D. Davidson and J. Hintikka, p.296.
15. "O.R.", p.35.
16. "O.R.", pp.35-38.
17. "O.R.", p.40.
18. "On the Reasons for Indeterminacy of Translation," op.cit.,p.183.
19. "O.R.", p.41.
20. "On the Reasons for Indeterminacy of Trans." op.cit., pp.181-82.
21. "O.R.", p.34.
22. Ibid.
23. "Identity and Predication," Journal of Philosophy, vol. 72, 1975, pp.343-64; hereafter referred to as "I.P."

Notes for Chapter 3 (cont.)

24. "I.P.", p.344.
25. D. Davidson, "Radical Interpretation," Dialectica, vol. 27, 1973, pp.313-28; D. Lewis, "Radical Interpretation," Synthese, vol. 27, 1974, pp.331-44.
26. R. Grandy, "Reference, Meaning and Belief," Journal of Philosophy, vol. 70, 1973, p.443.
27. "Radical Interpretation," op. cit., pp.336-39.
28. "I.P.", pp.345-46.
29. M. Dummett, "Truth," Proceedings of the Aristotelian Society, LIX, 1958-59, pp.141-62.
30. "I.P.", p.346.
31. Evans emphasises the importance of this point to the semantic theorist at the beginning of "I.P.", pp.343-44; see also p.346. More recently, however, Davidson has commented that though his point "still seems ... likely to be right,... [it] is not obvious." ("Radical Interpretation," op. cit., p.327, fn.2.)
32. "I.P.", p.346.
33. "I.P.", pp.350-53.
34. WQ, section 13.
35. "I.P.", p.351.
36. Ibid.
37. "I.P.", p.352.
38. "I.P.", p.353.
39. Ibid.
40. See in particular "I.P.", section III, paragraph 3.
41. "Existence and Quantification," Ontological Relativity and Other Essays, p.103.
42. Ibid.; see also WQ, section 13.
43. WQ, p.61.
44. WQ, p.52.
45. Ibid.
46. P. Strawson, "Singular Terms and Predication," Words and Objections, op. cit., p.100.
47. "Replies," Words and Objections, op. cit., pp.320-21.
48. "Singular Terms and Predication," op. cit., p.100.

Notes for Chapter 3 (cont.)

49. "Replies," op. cit., p.321.
50. "Existence and Quantification," op. cit., pp.104-5.
51. Ibid., p.105.
52. Ibid.
53. "Gavagai," Analysis, vol. 32, 1971-72, p.71.
54. Ibid.
55. "Comment on Donald Davidson," Synthese, vol. 27, 1974, pp.326-7.
56. "Gavagai," op. cit., p.71; Hill also presents formulae similar to my (1), (2) and (3) on p.106, but his support for them and the conclusions he draws from them differ markedly from what I say.
57. WO, p.52.
58. Ibid.

Notes for Chapter 4

1. Sophist, 261c-263d.
2. Metaphysics, 1011b26.
3. "The Concept of Truth in Formalized Languages," reprinted in A. Tarski, Logic, Semantics, Metamathematics; hereafter referred to as CTFL. My appreciation of the sense in which Tarski's theory is a correspondence theory has been helped by reading D. Davidson, "True to the Facts," Journal of Philosophy, vol. 66, 1969, pp.748-64.
4. The argument is taken further in J. Skorupski, Symbol and Theory, Appendix. Skorupski thinks the relativist can defend himself against the 'truth' objection I have outlined, but not against a 'meaning' objection similar to that urged against Feyerabend's position at the beginning of Chapter 2.
5. M. Dummett, Frege, Philosophy of Language, Chapter 13.
6. Ibid., p.444. It should be noted that Dummett himself thinks this condition is not met by correspondence theories, but this is because he assumes that the correspondence holds between a statement as a whole and something else. Later in this section I shall explain how truth can be defined in terms of a correspondence relation of a different sort.
7. Ibid., p.445.
8. Notice that this is quite different from the relativist's view that a sentence is true relative to some theoretical framework or set of core statements. The framework, or set of statements, is interpreted, and held true by those who share it. Once it is rejected there is no guarantee that any sentence will retain its original truth-value.
9. S. Kripke, "Outline of a Theory of Truth," Journal of Philosophy, vol. 72, 1975, pp.690-716.
10. In CTFL Tarski calls 'closed sentences' just 'sentences' and 'open sentences' 'sentential functions'.
11. H. Putnam, Meaning and the Moral Sciences, p.10.
12. W. Quine, Philosophy of Logic, p.40.
13. CTFL, pp. 152-53.
14. CTFL, p.194, fn.1.

Notes for Chapter 4 (cont.)

15. In "Is There a Problem about Substitutional Quantification?", Truth and Meaning, ed. G. Evans and J. McDowell, S. Kripke argues that, contrary to the claims of J. Wallace and L. Tharp, it is possible to meet Tarski's Convention T, and hence to define the concept of truth, using a substitutional interpretation of the quantifiers. If this were possible, my claim that Tarski's definition makes "essential use" of words/world relations might be contested on the ground that it implies that a referential interpretation is obligatory. How successful Kripke's attempt is remains to be seen. It depends on whether the metalanguage has to be counted referential. Quine has pointed out, in his review of Truth and Meaning (Journal of Philosophy, vol. 74, 1977, pp.225-41), that as it stands Kripke's truth definition violates Davidson's requirement that all T-sentences be deducible from a finite set of axioms. Quine makes a tentative suggestion (pp.237-38) for how "finitude might be regained," but clearly further work is required. If Kripke's definition can be made good, my claim that reference underwrites truth might need revising to "reference underwrites the definition of truth (for a language of type L) given by Tarski." But it will still not affect the main argument of this section, that the definition given by Tarski presupposes certain primitive relations of reference of a sort to be indicated.
16. "Tarski's Theory of Truth," Journal of Philosophy, 1972, pp.347-75.
17. *Ibid.*, pp.362-63.
18. *Ibid.*, p.363.
19. *Ibid.*, p.367.
20. *Ibid.*, p.373.
21. Meaning and the Moral Sciences, p.16.
22. "Tarski's Theory of Truth," *op. cit.*, p.355.
23. *Ibid.*
24. Meaning and the Moral Sciences, p.17.
25. "Tarski's Theory of Truth," *op. cit.*, p.368.
26. CTFL, pp.193-94; "The Semantic Conception of Truth and the Foundations of Semantics," Philosophy and Phen. Res., 1944, p.351.

Notes for Chapter 5

1. H. Putnam, "The Analytic and the Synthetic," reprinted in his Mind, Language and Reality, Philosophical Papers: Volume 2, p.52. He cites Wittgenstein's rope metaphor as another way of expressing what he understands by a cluster. It should be noted that Putnam has since come to think such an account is incorrect — see "Language and Reality," same volume, p.281. His reasons for doing so, however, seem to me to be misplaced. In particular, the weight he comes to place on the "Principle of Reasonable Ignorance" is already adequately borne by the "Principle of the Division of Linguistic Labour", for which see later in this section.
2. "The Analytic and the Synthetic," op. cit., p.53.
3. S. Kripke, "Naming and Necessity," (hereafter "N.N.") Semantics of Natural Language, ed. D. Davidson and G. Harman; H. Putnam, "The Meaning of 'Meaning'," reprinted in his Mind, Language and Reality, Philosophical Papers: Volume 2.
4. See, e.g., E. Zemach, "Putnam's Theory on the Reference of Substance Terms," Journal of Philosophy, vol. 73, 1976; M. Dummett, Frege, Philosophy of Language, Appendix to Ch. 5; D. H. Mellor, "Natural Kinds," unpublished ms., Cambridge.
5. F. Cajori, A History of Physics, p.359; I owe this reference to A. Fine, "How to Compare Theories: Reference and Change," Nous, vol. 9, 1975.
6. "On Sense and Reference," Philosophical Writings of Gottlob Frege, trans. and ed. P. Geach and M. Black, p.58, fn.
7. M. Dummett, Frege, Philosophy of Language, p.110.
8. "The Meaning of 'Meaning'," op. cit., p.228.
9. L. Wittgenstein, Philosophical Investigations, para. 79.
10. J. Searle, "Proper Names," reprinted in Philosophical Logic, ed. P. Strawson.
11. P. Strawson, Individuals, pp.191-92.
12. Contained in Semantics of Natural Language, op. cit.
13. "N.N." pp.292-93; also G. Evans, "The Causal Theory of Names," Aristotelian Society, Supplementary Volume 47, 1973, pp.187-89.
14. "N.N." Lecture I, especially pp.274ff.

Notes for Chapter 5 (cont.)

15. Frege, Philosophy of Language, Appendix to Ch.5.
16. Ibid., p.131; the argument for this conclusion begins on p.111.
17. Ibid.
18. Cf. *ibid.*, p.132. I have altered Dummett's example in two respects: firstly, he is not concerned so much with a cluster theory but with a Fregean descriptive theory; secondly, his example, chosen to accord with Kripke's criticism, concerns a name whose sense is supposed to be given by a single definite description.
19. "N.N." p.294.
20. "N.N." p.348, fn.36.
21. "N.N." p.279.
22. "N.N." p.276 and Lecture III.
23. Ibid., see particularly pp.350-51, fn's.56,57.
24. "N.N." *passim*; for criticism of the notion see M. Dummett, "Postscript," Synthese, vol. 27, 1974, and Frege, Philosophy of Language, pp.116-20.
25. Cf. "N.N." pp.301-2.
26. "N.N." p.302.
27. Frege, Philosophy of Language, pp.146-51.
28. Ibid., p.149.
29. "The Causal Theory of Names," *op. cit.*
30. Frege, Philosophy of Language, p.150.
31. Kripke makes some brief remarks about the 'Madagascar' example in the Addenda to "N.N.", Semantics of Natural Language, *op.cit.*, pp.768-69, but they completely misunderstand the point Evans and Dummett make.
32. Frege, Philosophy of Language, p.151.
33. "The Causal Theory of Names," *op.cit.*, p.197.
34. Ibid.
35. W. Quine, Word and Object, Ch. II.
36. H. Putnam, "Comment on Wilfrid Sellars," Synthese, vol. 27, 1974, p.449.
37. "Postscript," *op.cit.*, pp.530-31.
38. Ibid., p.531.

Notes for Chapter 5 (cont.)

39. "N.N.", Lecture III; H. Putnam, "It Ain't Necessarily So," reprinted in his Mathematics, Matter and Method, Philosophical Papers: Volume 1, "The Meaning of 'Meaning'," op.cit., and "Comment on Wilfrid Sellars," op.cit. The examples Putnam gives in part I of "It Ain't Necessarily So" are much the same as some of Kripke's and so will not be discussed separately.
40. "N.N." p.318.
41. "Postscript," op.cit., p.532.
42. "Comment on Wilfrid Sellars," op.cit., p.447. In "The Meaning of 'Meaning'," op.cit., the first thesis is sometimes phrased in terms of a speaker's psychological state.
43. These are the elm/beech one and the aluminium/molybdenum one; see "Comment on Wilfrid Sellars," op.cit., p.450, and "The Meaning of 'Meaning'," op.cit., pp.225-27.
44. Cf. Putnam's exposition in "Comment on Wilfrid Sellars," op.cit., p.451, and "The Meaning of 'Meaning'," op.cit., pp.223-25.
45. "Putnam's Theory on the Reference of Substance Terms," op.cit., p.120.
46. "Comment on Wilfrid Sellars," op.cit., p.451; quotation marks added.
47. Ibid., cf. "N.N." Lecture I.
48. "N.N." p.320.
49. "Putnam's Theory on the Reference of Substance Terms," op.cit., pp.121-22.
50. "N.N." p.320, my emphasis.

Notes for Chapter 6

1. W. Quine, Word and Object, p.27.
2. Ibid.
3. Ibid., p.73.
4. W. Quine, "On the Reasons for Indeterminacy of Translation," Journal of Philosophy, vol. 67, 1970, pp.179-80.
5. Ibid.
6. W. Quine, "Reply to Chomsky," Words and Objections, ed. D. Davidson and J. Hintikka; p.303.
7. "On the Reasons for Indeterminacy of Translation," op. cit., p.182.
8. Ibid.
9. Ibid.
10. Ibid., and references there.
11. The locus classicus for discussion of underdetermination is of course P. Duhem, The Aim and Structure of Physical Theory, particularly Ch. VI. Further examples and discussion are to be found in W. Newton-Smith, "The Underdetermination of Theory by Datum," Aristotelian Society, Supplementary Volume 52, 1978.
12. C. Hookway, "Indeterminacy and Interpretation," Action and Interpretation, ed. C. Hookway and P. Pettit, p.24.
13. D. Lewis, "Radical Interpretation," Synthese 27, 1974, pp.331-44; for critical discussion see D. Davidson, "Replies to Lewis and Quine," ibid., pp.345-7.
14. R. Grandy, "Reference, Meaning and Belief," Journal of Philosophy, 1973.
15. C. Hookway, op. cit., p.27.
16. Word and Object, p.59.
17. N. Wilson, "Substances Without Substrata," Review of Metaphysics 12, 1959, pp.521-39; quoted by Quine, Word and Object, p.59, fn.2.
18. W. Quine, Ontological Relativity and Other Essays, p.46.
19. His various papers on this theme are "Radical Interpretation," Dialectica 27, 1973, pp.313-28; "On the Very Idea of a Conceptual Scheme," Proceedings of the American Philosophical Association 47, 1973-4, pp.5-20; "In Defence of Convention T," Truth, Syntax and Modality, ed. H. Leblanc, pp.76-86; "Belief and the Basis of

Notes for Chapter 6 (cont.)

- Meaning," *Synthese* 27, 1974, pp.309-23; "Thought and Talk," *Mind and Language*, ed. S. Guttenplan, pp.7-23.
20. "Belief and the Basis of Meaning," op. cit., p.321.
21. "On the Very Idea of a Conceptual Scheme," op. cit., p.19.
22. Ibid.
23. Ibid.
24. C. McGinn, "Charity, Interpretation, and Belief," Journal of Philosophy, vol. 74, 1977, p.523.
25. "Radical Interpretation," p.324.
26. "Charity, Interpretation, and Belief," op. cit., p.524; he quotes from "Thought and Talk," op. cit., pp. 20-21.
27. "Charity, Interpretation, and Belief," op. cit., p.526.
28. "Thought and Talk," op. cit., pp.20-21.
29. "Belief and the Basis of Meaning," op. cit., p.318.
30. "Radical Interpretation," op. cit., p.323.
31. "Belief and the Basis of Meaning," op. cit., pp.319-20.
32. "Radical Interpretation," op. cit., p.324.
33. W. Quine, "Comment on Donald Davidson," *Synthese* 27, 1974, pp.325-29.
34. Ibid., p.326.
35. Ibid.
36. Ibid., p.328.
37. "In Defence of Convention T," op. cit., p.84; see also "Replies to Lewis and Quine," op. cit., pp.347-8.
38. Reprinted in W. Quine, From a Logical Point of View.
39. See, e.g., Word and Object, p.42. This is perhaps being too charitable to Quine. It has been argued by C. Boorse, "The Origins of the Indeterminacy Thesis," Journal of Philosophy, vol. 72, 1975, that all of Quine's arguments for indeterminacy presuppose a clear observation/theoretical distinction. If this is the case then so much the worse for Quine. It would then seem that the radical translator would have one less constraint, but the radical interpreter, since he makes no use of the Quinean notion of "stimulus meaning", will not be affected.

Notes for Chapter 6 (cont.)

40. J. Dalton, A New System of Chemical Philosophy, p.163.
41. K. Parsons, "A Criterion for Meaning Change," Philosophical Studies, 28, 1975, p.379.
42. I. Hacking, Why Does Language Matter to Philosophy?, p.154.
43. Ibid.
44. "Radical Interpretation," op. cit., p.321.
45. "In Defence of Convention T," op. cit., p.84, my emphasis.
46. "On the Reasons for Indeterminacy of Translation," pp.179-80.
47. "Experiment in Plant-hybridization," Verh. naturf. Ver. in Brunn, Abh. iv, 1865.
48. Reprinted in Mind, Language and Reality, Philosophical Papers: Vol. 2; see section IIA.
49. Ibid., p.205.
50. H. Putnam, Meaning and the Moral Sciences, p.58.
51. "Reply to Chomsky," op. cit., p.303.
52. Ontological Relativity and Other Essays, pp.26-7.
53. D. Føllesdal, "Meaning and Experience," Mind and Language, ed. S. Guttenplan, p.32; the argument is discussed at greater length in his "Indeterminacy of Translation and Underdetermination of the Theory of Nature," Dialectica, 27, 1973.
54. "Meaning and Experience," op. cit., p.32.
55. Ontological Relativity and Other Essays, p.26.

Notes for Chapter 7

1. A jibe Jardine suggests a relativist might direct at a realist account of the growth of science, "'Realistic' Realism and the Progress of Science," Action and Interpretation, ed. C. Hookway and P. Pettit, p.107.
2. "'Realistic' Realism and the Progress of Science," op. cit., particularly section IV.
3. Putnam makes extensive use of the principle in "Language and Reality," reprinted in his Mind, Language and Reality, Philosophical Papers: Volume 2, and in Meaning and the Moral Sciences.
4. Meaning and the Moral Sciences, p.22.
5. "'Realistic' Realism and the Progress of Science," op.cit, p.112.
6. Ibid., p.122.
7. Ibid., p.123.
8. One reason for thinking so is given towards the end of the section. In general it would seem that at times of "crisis" (in Kuhn's sense), when there are frequent contradictions between theory and observation, we might often be obliged to assign a null extension.
9. P. Feyerabend, "Problems of Empiricism," Frontiers of Science and Philosophy, ed. R. Colodny, pp.168ff; cf. my Chap. 1, sec.(iv).
10. "The Analytic and the Synthetic," reprinted in his Mind, Meaning and Language, Philosophical Papers: Volume 2, p.52.
11. Ibid.
12. Ibid.
13. In passing it might be noted that all of Feyerabend's many examples involve change of this kind.
14. Journal of Philosophy, vol. 70, 1973, pp.462-81.
15. Ibid., p.462.
16. "Against Indeterminacy," Journal of Philosophy, vol. 74, 1977.
17. Ibid., p.535. It is explained in J. Earman and M. Friedman, "The Meaning and Status of Newton's Law of Inertia," Philosophy of Science, vol. 40, 1973, pp.329-59.
18. "Against Indeterminacy," op. cit., p.536.
19. Ibid.
20. "Problems of Empiricism," op. cit., p.169.